目　　录

第一章　预 备 知 识

§1　特征值与特征向量

本节列举矩阵代数的几条熟知的定义和结论.

定义 1.1. 设 $A \in \mathbf{C}^{n\times n}$. 如果存在 $\lambda \in \mathbf{C}$ 和非零向量 $\boldsymbol{x} \in \mathbf{C}^n$, 使得 $A\boldsymbol{x} = \lambda\boldsymbol{x}$, 则 λ 叫做 A 的特征值, $\boldsymbol{x}$ 叫做 A 的属于特征值 λ 的特征向量.

A 的所有特征值的全体,叫做 A 的谱 (spectrum), 记作 $\lambda(A)$.

据定义 1.1, $\lambda \in \lambda(A)$ 的必要与充分条件是

$$\det(\lambda I - A) = 0.$$

$p(\lambda) \equiv \det(\lambda I - A)$ 叫做 A 的特征多项式. 如果 A 有 r 个不同的特征值 $\lambda_1, \cdots, \lambda_r$, 其重数分别为 $n(\lambda_1), \cdots, n(\lambda_r)$, 则

$$p(\lambda) = \prod_{i=1}^{r} (\lambda - \lambda_i)^{n(\lambda_i)}, \quad \lambda_i \neq \lambda_j \ (i \neq j),$$

$$\sum_{i=1}^{r} n(\lambda_i) = n,$$

其中 $n(\lambda_i)$ 叫做 λ_i 的代数重数.如果

$$\operatorname{rank}(\lambda_i I - A) = n - m(\lambda_i),$$

则 $m(\lambda_i)$ 叫做 λ_i 的几何重数, 它表示 A 的属于 λ_i 的线性无关特征向量的个数. 显然有

$$1 \leqslant m(\lambda_i) \leqslant n(\lambda_i) \leqslant n.$$

定理 1.1. 设 $A \in \mathbf{C}^{n\times n}$ 有 r 个不同的特征值 $\lambda_1, \cdots, \lambda_r$, 其代数重数分别为 $n(\lambda_1), \cdots, n(\lambda_r)$, 则必存在非奇异阵 $P \in \mathbf{C}^{n\times n}$, 使得

$$P^{-1}AP = J \equiv \operatorname{diag}(J_1(\lambda_1), \cdots, J_r(\lambda_r)), \tag{1.1}$$

其中

$$J_i(\lambda_i)=\mathrm{diag}(J_i^{(1)}(\lambda_i),\cdots,J_i^{(k_i)}(\lambda_i))\in \mathbf{C}^{n(\lambda_i)\times n(\lambda_i)},\ 1\leqslant i\leqslant r, \tag{1.2}$$

$$J_i^{(k)}(\lambda_i)=\begin{pmatrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i\end{pmatrix}\in \mathbf{C}^{n_k(\lambda_i)\times n_k(\lambda_i)},1\leqslant k\leqslant k_i, \tag{1.3}$$

$$\sum_{k=1}^{k_i} n_k(\lambda_i)=n(\lambda_i),\ 1\leqslant i\leqslant r;$$

并且除了 $J_i^{(k)}(\lambda_i)$ ($1\leqslant k\leqslant k_i$, $1\leqslant i\leqslant r$) 的编排次序可以改变外，J 是唯一确定的.

(1.3) 式所示的每个矩阵 $J_i^{(k)}(\lambda_i)$ ($1\leqslant k\leqslant k_i$, $1\leqslant i\leqslant r$) 叫做 Jordan 块;矩阵 J 叫做 A 的 Jordan 标准形.

定理 1.1 是矩阵论的基本结果之一，它的证明可以在普通的线性代数教程中找到.

定义 1.2. 设 $A\in \mathbf{C}^{n\times n}$. 如果存在非奇异阵 $Q\in \mathbf{C}^{n\times n}$,使得

$$Q^{-1}AQ=\mathrm{diag}(\lambda_1,\cdots,\lambda_n),$$

则 A 叫做可对角化阵(又叫单构阵,或者可正规化阵).

易知下面 5 种叙述互相等价:

1) A 是可对角化阵，

2) $\mathbf{C}^n$ 存在由 A 的特征向量构成的一组基底，

3) A 的初等因子都是线性的，

4) A 的 Jordan 标准形中的 Jordan 块都是 1 阶的，

5) $m(\lambda_i)=n(\lambda_i)$, $\forall\lambda_i\in\lambda(A)$.

定义 1.3. 设 $A\in \mathbf{C}^{n\times n}$. 如果 $AA^H=A^HA$, 则称 A 为正规矩阵（以下简称之为正规阵）; 如果 $A^H=A$, 则称 A 为 Hermite 阵;如果 $A^T=\bar{A}=A$,则称 A 为实对称阵;如果 $A^HA=I$, 则称 A 为酉阵;如果 $A^TA=I$ 并且 $\bar{A}=A$, 则称 A 为实正交阵.

下面的定理 1.2 是熟知的(§ 2 习题 4 提出了一种证法).

定理 1.2 (Schur 定理). 设 $A\in \mathbf{C}^{n\times n}$, 则存在酉阵 U, 使得

$$U^HAU=T, \tag{1.4}$$

其中T是上三角阵；而且适当选取U，可使T的对角线元素按任一指定顺序排列。

(1.4)式右端的T，叫做A的 Schur 上三角形式。可记

$$T = \Lambda + M, \quad \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

其中M为严格上三角阵(即对角线元素为零的上三角阵)。

从定理 1.2 可得

推论 1.1. 下列结论成立:

1) A是正规阵$\Longleftrightarrow$存在酉阵U，使得

$$U^H A U = \mathrm{diag}(\lambda_1, \cdots, \lambda_n),$$

即$\mathbf{C}^n$存在由A的特征向量构成的标准正交基。

2) A是 Hermite 阵$\Longleftrightarrow A$是正规阵，并且$\lambda(A) \subseteq \mathbf{R}$.

3) A是酉阵$\Longleftrightarrow A$是正规阵，并且

$$\lambda(A) \subseteq \mathscr{S} = \{\lambda \in \mathbf{C} : |\lambda| = 1\}.$$

同理可得实对称阵和实正交阵的相类似的结论。

定义 1.4. 设 $A \in \mathbf{C}^{n \times n}$ 是 Hermite 阵。如果

$$\boldsymbol{x}^H A \boldsymbol{x} > 0, \ \forall \boldsymbol{x} \in \mathbf{C}^n, \ \boldsymbol{x} \neq 0$$

(或 $\boldsymbol{x}^H A \boldsymbol{x} \geqslant 0, \ \forall \boldsymbol{x} \in \mathbf{C}^n$)，则$A$叫做正定(或半正定)阵。

以下用 $A > 0$ ($A \geqslant 0$) 表示A是正定(半正定)阵。

设 $A \in \mathbf{C}^{n \times n}$ 为 Hermite 阵。容易验证下面的5种叙述是互相等价的:

1) A是正定(半正定)阵，

2) A的每个特征值$>0(\geq 0)$，

3) A的每个主子阵都是正定(半正定)阵，

4) A的每个主子式(即主子阵的行列式)$>0(\geq 0)$，

5) 对任一 $n \times n$ 非奇异阵 Q，$Q^H A Q$ 是正定(半正定)阵。

习题

1. 设A是正规阵，同时A是上三角阵。证明A是对角阵。

2. 设A为可对角化阵。试证：A的特征值皆为实数的必要与充分条件是存在正定阵H，使得 HA 为 Hermite 阵。

3. 设 $A\in \mathbf{C}^{n\times n}$，$A$ 的所有特征值互不相等．则 A 的特征向量构成 $\mathbf{C}^n$ 的一组基底．

4. 设 $AB=BA$，其中 B 为幂零阵（即存在自然数 m，使得 $B^m=0$）．则 $\det(A+B)=\det A$．

5. 设 $A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$，其中 A_{ij} 均为方阵，并且满足 $A_{11}A_{21}=A_{21}A_{11}$．则

$$\det A=\det(A_{11}A_{22}-A_{21}A_{12}).$$

6. 设 $\det\begin{pmatrix} A & \boldsymbol{a} \\ \boldsymbol{a}^H & \alpha \end{pmatrix}=0$．则

$$\det\begin{pmatrix} A & \boldsymbol{a} \\ \boldsymbol{a}^H & \beta \end{pmatrix}=(\beta-\alpha)\det A.$$

7. 设 A 与 B 均为可对角化阵．试证下列三个条件互相等价：

i) $AB=BA$，

ii) 存在非奇异阵 X，使 $X^{-1}AX$ 与 $X^{-1}BX$ 同时为对角阵，

iii) 存在一个可对角化阵 C 及一对多项式 $p(\lambda)$ 与 $q(\lambda)$，使得 $A=p(C)$ 和 $B=q(C)$．

8. 至少给出两种不同的方法，计算行列式 $\det(I-\sigma\boldsymbol{u}\boldsymbol{v}^H)$，其中 $\boldsymbol{u},\boldsymbol{v}\in\mathbf{C}^n$，$\sigma\in\mathbf{C}$．

9. 证明 AB 与 BA 的特征多项式除了一个 λ 的幂之外是相等的．

10. 设 H_1 为正定阵，H_2 是与 H_1 同阶的 Hermite 阵．试证：H_1+H_2 为正定阵的必要与充分条件是 $H_1^{-1}H_2$ 的特征值均大于 -1．

§2 初等矩阵

2.1 初等矩阵的一般形式

定义 2.1. 形如

$$E(\boldsymbol{u},\boldsymbol{v};\sigma)=I-\sigma\boldsymbol{u}\boldsymbol{v}^H$$

的矩阵叫做初等矩阵，其中 $u,v\in\mathbb{C}^n$，$\sigma\in\mathbb{C}$.

初等矩阵是数值代数的基本工具之一。

显然 $E(u,v;0)=I$. 所以下面仅讨论 $\sigma\neq 0$ 的情形.

初等矩阵的性质:

1）特征向量. 若 $u\notin v^{\perp}$（$v^{\perp}$ 表示与 v 正交的 $n-1$ 维子空间)，则 $E(u,v;\sigma)$ 有 n 个线性无关的特征向量，该组特征向量由 u 及 $v^{\perp}$ 中任取一组基底构成；若 $u\in v^{\perp}$，则$E(u,v;\sigma)$仅有 $n-1$ 个线性无关的特征向量，该组特征向量由 $v^{\perp}$ 中任取一组基底构成.

证明:

首先在 $v^{\perp}$ 中任取一组基底 $\{u_i\}_{i=1}^{n-1}$，有

$$E(u,v;\sigma)\ u_i=u_i,\ \ i=1,\cdots,n-1.$$

因此 $u_i(i=1,\cdots,n-1)$是 $E(u,v;\sigma)$ 的属于特征值 1 的特征向量.

若 $u\notin v^{\perp}$，则有

$$E(u,v;\sigma)\ \ u=(1-\sigma v^H u)u,$$

即 u 也是 $E(u,v;\sigma)$ 的特征向量，相应的特征值是 $1-\sigma v^H u$.

若 $u\in v^{\perp}$，则这时 $E(u,v;\sigma)$ 的任一特征向量 x 必有 $x\in v^{\perp}$. 事实上，由

$$E(u,v;\sigma)x=qx$$

可得

$$(1-q)x=\sigma v^H x u.$$

如果 $q=1$，则有 $v^H x=0$，即 $x\in v^{\perp}$；如果 $q\neq 1$，假定 $x\notin v^{\perp}$，则 x 应与 u 共线，但这与已知 $u\in v^{\perp}$ 矛盾，所以这时亦有 $x\in v^{\perp}$.

2）特征值. 在 $v^{\perp}$ 内取一标准正交基 $u_2,\cdots,u_n$，构造酉阵

$$U=\left(\frac{v}{\|v\|_2},\ u_2,\cdots,u_n\right).$$

由

$$uv^H U = U\begin{pmatrix} v^H u & & & \\ * & 0 & & \\ \vdots & & \ddots & \\ * & & & 0 \end{pmatrix}$$

可得

$$E(u, v; \sigma)U = U\left(I - \sigma\begin{pmatrix} v^H u & & & \\ * & 0 & & \\ \vdots & & \ddots & \\ * & & & 0 \end{pmatrix}\right),$$

即

$$U^H E(u,v;\sigma)U = \begin{pmatrix} 1-\sigma v^H u & & & \\ * & 1 & & \\ \vdots & & \ddots & \\ * & & & 1 \end{pmatrix}.$$

因此

$$\lambda(E(u,v;\sigma)) = \{1-\sigma v^H u, 1, \cdots, 1\}.$$

3) 行列式. 由 2)立即得出

$$\det E(u,v;\sigma) = 1 - \sigma v^H u.$$

4) 逆矩阵. 由 3)及

$$E(u,v;\sigma)E(u,v;\tau) = E(u,v;\sigma+\tau-\sigma\tau v^H u)$$

可知,当且仅当 $\sigma v^H u \neq 1$ 时, $E(u,v;\sigma)$ 可逆,并且

$$E(u,v;\sigma)^{-1} = E\left(u, v; \frac{\sigma}{\sigma v^H u - 1}\right).$$

特别地,当 $\sigma v^H u = 0$ 时,有

$$E(u,v;\sigma)^{-1} = E(u,v;-\sigma);$$

当 $\sigma v^H u = 2$ 时,有

$$E(u,v;\sigma)^{-1} = E(u,v;\sigma).$$

5) 对于任意非零向量 a 与 $b \in \mathbb{C}^n$, 必可适当选取 u, v 与 σ, 使得

$$E(u,v;\sigma)a = b. \tag{2.1}$$

事实上,只需取 u, v 与 σ 满足

$$\boldsymbol{v}^H\boldsymbol{a} \neq 0, \quad \sigma\boldsymbol{u} = \frac{\boldsymbol{a}-\boldsymbol{b}}{\boldsymbol{v}^H\boldsymbol{a}} \tag{2.2}$$

即可.

(2.2)式表明,任给$\boldsymbol{a}$与$\boldsymbol{b} \in \mathbf{C}^n$,为了使(2.1)式成立,$\boldsymbol{u}$, $\boldsymbol{v}$ 与σ在选取上有很大的灵活性. 因此,可以选取各种特殊的$\boldsymbol{u}$,$\boldsymbol{v}$ 与σ,以满足各种不同的要求.

附带指出,所有的初等变换矩阵,都可以表示成 $E(\boldsymbol{u},\boldsymbol{v};\sigma)$ 的形式:

(i) 交换第 i, j 两行,即在矩阵的左边乘上

$$I_{i,j} \equiv I - (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^T = E(\boldsymbol{e}_i - \boldsymbol{e}_j, \boldsymbol{e}_i - \boldsymbol{e}_j; 1);$$

(ii) 第 i 行乘以 α 加到第 j 行,即在矩阵的左边乘上

$$I + \alpha\boldsymbol{e}_j\boldsymbol{e}_i^T = E(\boldsymbol{e}_j, \boldsymbol{e}_i; -\alpha);$$

(iii) 第 i 行乘以 α, 即在矩阵的左边乘上

$$I - (1-\alpha)\boldsymbol{e}_i\boldsymbol{e}_i^T = E(\boldsymbol{e}_i, \boldsymbol{e}_i; 1-\alpha).$$

其中 $\boldsymbol{e}_i$ 表示单位阵 I 的第 i 列.

下面分别说明两种不同的初等矩阵.

2.2 初等下三角阵

令

$$\boldsymbol{l}_i = (0, \cdots, 0, l_{i+1,i}, \cdots, l_{ni})^T, \tag{2.3}$$

则

$$L_i(\boldsymbol{l}_i) \equiv E(\boldsymbol{l}_i, \boldsymbol{e}_i; 1)$$

叫做初等下三角阵.即

$$L_i(\boldsymbol{l}_i) = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & 0 & \\ & & 1 & & & \\ & & -l_{i+1,i} & 1 & & \\ & & \vdots & & \ddots & \\ & & -l_{n,i} & & & 1 \end{pmatrix} \begin{matrix} \\ \\ i \\ i+1 \\ \\ \\ \end{matrix}$$

$$\qquad\qquad i$$

用 $L_i(l_i)$ 左乘一个矩阵 A,等于从A的第 k 行减去第 i 行乘以 l_{ki} ($k = i+1, \cdots, n$). 对于 $A = (a_{ij})$, 如果取

$$l_{ki} = \frac{\alpha_{ks}}{\alpha_{is}} \quad (k = i + 1, \cdots, n),$$

则 $L_i(\boldsymbol{l}_i)A$ 的第 $(i+1,s),\cdots,(n,s)$ 元素为零. 这就是消去法的一步.

易知下列关系式成立:

$$E(\boldsymbol{l}_i, \boldsymbol{e}_i; 1)^{-1} = E(\boldsymbol{l}_i, \boldsymbol{e}_i; -1); \tag{2.4}$$

$$E(\boldsymbol{l}_i, \boldsymbol{e}_i; 1)E(\boldsymbol{l}_j, \boldsymbol{e}_j; 1) = I - \boldsymbol{l}_i\boldsymbol{e}_i^T - \boldsymbol{l}_j\boldsymbol{e}_j^T \ (j \geqslant i). \tag{2.5}$$

注意: $E(\boldsymbol{l}_i, \boldsymbol{e}_i; 1)$与 $E(\boldsymbol{l}_j, \boldsymbol{e}_j; 1)$ 的乘法，一般不可交换.

由(2.5)可知,任一单位下三角阵

$$L = \begin{pmatrix} 1 & & & \\ -l_{21} & 1 & & \\ \vdots & \ddots & \ddots & \\ -l_{n1} & \cdots & -l_{n,n-1} & 1 \end{pmatrix}$$

必可分解为

$$L = L_1(\boldsymbol{l}_1)L_2(\boldsymbol{l}_2)\cdots L_{n-1}(\boldsymbol{l}_{n-1}), \tag{2.6}$$

其中 $\boldsymbol{l}_i$ 如(2.3)所示. 事实上,(2.6)式的右端等于

$$\begin{aligned}
&(I - \boldsymbol{l}_1\boldsymbol{e}_1^T)(I - \boldsymbol{l}_2\boldsymbol{e}_2^T)\cdots(I - \boldsymbol{l}_{n-1}\boldsymbol{e}_{n-1}^T) \\
&\quad = (I - \boldsymbol{l}_1\boldsymbol{e}_1^T - \boldsymbol{l}_2\boldsymbol{e}_2^T)(I - \boldsymbol{l}_3\boldsymbol{e}_3^T)\cdots(I - \boldsymbol{l}_{n-1}\boldsymbol{e}_{n-1}^T) \\
&\quad = \cdots \\
&\quad = I - \boldsymbol{l}_1\boldsymbol{e}_1^T - \boldsymbol{l}_2\boldsymbol{e}_2^T - \cdots - \boldsymbol{l}_{n-1}\boldsymbol{e}_{n-1}^T = L.
\end{aligned}$$

因此，由(2.6)可以进一步得出: 任一非奇异下三角阵 L，必可表示成一个非奇异对角阵和若干个初等下三角阵的乘积.

2.3 初等 Hermite 阵

初等 Hermite 阵是指

$$H(\boldsymbol{w}) \equiv E(\boldsymbol{w}, \boldsymbol{w}; 2), \quad \boldsymbol{w}^H\boldsymbol{w} = 1.$$

$H(\boldsymbol{w})$ 又叫做初等酉阵,或 Householder 变换.

$H(\boldsymbol{w})$ 的性质:

1) $H(\boldsymbol{w})^H = H(\boldsymbol{w}) = H(\boldsymbol{w})^{-1}$.

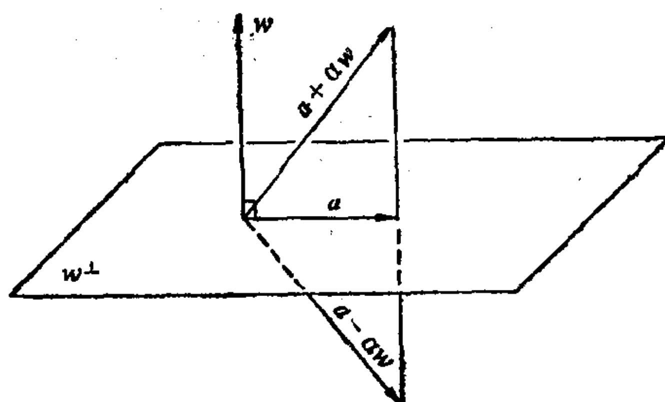


图 1—1

2) $\det H(\boldsymbol{w}) = -1$, $H(\boldsymbol{w}) = W\begin{pmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}W^H$,

其中 $W = (\boldsymbol{w}, W_1)$ 为酉阵.

3) $H(\boldsymbol{w})$ 是镜象变换. 具体地说,对于任一 $\boldsymbol{a} \in \boldsymbol{w}^\perp$, 有

$$H(\boldsymbol{w})(\boldsymbol{a} + \alpha\boldsymbol{w}) = \boldsymbol{a} - \alpha\boldsymbol{w},\ \alpha \in \mathbf{C}.$$

即 $H(\boldsymbol{w})$ 是关于 $\boldsymbol{w}$ 的垂直超平面的反演,如图 1-1 所示

4) 设 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 是 $\mathbf{C}^n$ 中的向量, 则存在单位向量 $\boldsymbol{w}$, 使得 $H(\boldsymbol{w})\boldsymbol{a} = \boldsymbol{b}$ 的必要与充分条件是

$$\|\boldsymbol{a}\|_2 = \|\boldsymbol{b}\|_2 \text{ 和 } \boldsymbol{a}^H\boldsymbol{b} = \boldsymbol{b}^H\boldsymbol{a}; \tag{2.7}$$

并且在(2.7)式所示的条件下,使得 $H(\boldsymbol{w})\boldsymbol{a} = \boldsymbol{b}$ 成立的单位向量 $\boldsymbol{w}$ 可取为

$$\boldsymbol{w} = e^{i\theta}(\boldsymbol{a} - \boldsymbol{b})/\|\boldsymbol{a} - \boldsymbol{b}\|_2, \tag{2.8}$$

其中 θ 为任一实数.

证明:

条件(2.7)的必要性是显然的. 以下证明其充分性.

据 3), 可取一与 $\boldsymbol{a} - \boldsymbol{b}$ 共线的单位向量 $\boldsymbol{w}$, 如(2.8)式所示. 这时有

$$H(\boldsymbol{w})\boldsymbol{a} = \left(I - 2\,\frac{(\boldsymbol{a} - \boldsymbol{b})(\boldsymbol{a} - \boldsymbol{b})^H}{\|\boldsymbol{a} - \boldsymbol{b}\|_2^2}\right)\boldsymbol{a}$$

$$= \boldsymbol{a} - \frac{2(\boldsymbol{a}^H\boldsymbol{a} - \boldsymbol{b}^H\boldsymbol{b})(\boldsymbol{a} - \boldsymbol{b})}{\boldsymbol{a}^H\boldsymbol{a} + \boldsymbol{b}^H\boldsymbol{b} - \boldsymbol{a}^H\boldsymbol{b} - \boldsymbol{b}^H\boldsymbol{a}}.$$

利用条件(2.7)即可得到 $H(\boldsymbol{w})\boldsymbol{a} = \boldsymbol{b}$. □

注 2.1. 由 $H(\boldsymbol{w})$ 的性质4)可知,对于

$$\boldsymbol{a} = (\alpha_1, \cdots, \alpha_n)^T \neq 0,$$

如果令

$$p = \begin{cases} \|\boldsymbol{a}\|_2 & \text{当 } \alpha_1 = 0 \text{ 时} \\ e^{i\arg\alpha_1}\|\boldsymbol{a}\|_2 & \text{当 } \alpha_1 \neq 0 \text{ 时}, \end{cases}$$

并取

$$\boldsymbol{w} = (\boldsymbol{a} - p\boldsymbol{e}_1)/\|\boldsymbol{a} - p\boldsymbol{e}_1\|_2,$$

则有 $H(\boldsymbol{w})\boldsymbol{a} = p\boldsymbol{e}_1$.

注 2.2. 任一旋转必可分解为两个初等 Hermite 阵的乘积. 事实上,有

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \left[I - 2\begin{pmatrix} 0 \\ 1 \end{pmatrix}(0, 1) \right]$$

$$\times \left[I - 2\begin{pmatrix} \sin\dfrac{\theta}{2} \\ -\cos\dfrac{\theta}{2} \end{pmatrix}\left(\sin\frac{\theta}{2}, -\cos\frac{\theta}{2}\right) \right].$$

注 2.3. 上面 2.2 与 2.3 所述的是两类特殊的初等矩阵,它们在数值分析中最常使用. $L_i(\boldsymbol{l}_i)$ 简单,运算量小,但数值稳定性较差;$H(\boldsymbol{w})$ 具有正交性,数值稳定性好,但运算量大. 因此,是否能够找到运算量小和稳定性好的初等矩阵,是值得研究的问题(参看[35]).

习题

1. 设 $\boldsymbol{u}, \boldsymbol{v} \in \mathbf{C}^n$, 满足 $\boldsymbol{v}^H\boldsymbol{u} = 1$. 令 $S = I - 2\boldsymbol{u}\boldsymbol{v}^H$ 和 $\boldsymbol{b} = -2\beta\boldsymbol{u}$, $\beta \in \mathbf{C}$. 试证:

1) $S^{-1} = S$, $S\boldsymbol{b} = -\boldsymbol{b}$;

2) 若 $\boldsymbol{y} = S\boldsymbol{x} + \boldsymbol{b}$, 则 $\dfrac{\boldsymbol{x} + \boldsymbol{y}}{2}$ 属于超平面 $\boldsymbol{v}^H\boldsymbol{z} + \beta = 0$,

并且 $\boldsymbol{y}-\boldsymbol{x}$ 与 $\boldsymbol{u}$ 共线；

3）试给出一图示，说明 $\boldsymbol{y}=S\boldsymbol{x}+\boldsymbol{b}$ 的几何意义(注：$\boldsymbol{y}$ 叫做 $\boldsymbol{x}$ 沿方向 $\boldsymbol{u}$ 关于超平面 $\boldsymbol{v}^H\boldsymbol{z}+\beta=0$ 的斜镜象映射).

2. 利用初等矩阵的性质，证明：若 A 可逆，$\sigma\neq 0$，并且 $\boldsymbol{v}^HA^{-1}\boldsymbol{u}\neq\sigma^{-1}$，则

$$(A-\sigma\boldsymbol{u}\boldsymbol{v}^H)^{-1}=A^{-1}-\frac{A^{-1}\boldsymbol{u}\boldsymbol{v}^HA^{-1}}{\boldsymbol{v}^HA^{-1}\boldsymbol{u}-\sigma^{-1}}.$$

3. 设 $A\in\mathbf{C}^{n\times n}$, U，$V\in\mathbf{C}^{n\times m}(n\geqslant m)$，$S\in\mathbf{C}^{m\times m}$. 试证：若 A 与 S 均为非奇异阵，并且 $V^HA^{-1}U-S^{-1}$ 亦非奇异，则

$$(A-USV^H)^{-1}=A^{-1}-A^{-1}U(V^HA^{-1}U-S^{-1})^{-1}V^HA^{-1}.$$

4. 设 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，$\boldsymbol{x}\neq 0$. 试证：存在初等 Hermite 阵 H，使得 $HAH\boldsymbol{e}_1=\lambda\boldsymbol{e}_1$. 并由此证明 Schur 定理(即定理 1.2).

5. 设 $T=(\tau_{ij})$ 为上三角阵. 如果对于某二指标 i 与 $j(i<j)$，$\tau_{ii}\neq\tau_{jj}$，则可适当选取 σ，使得

$$\boldsymbol{e}_i^TE(\boldsymbol{e}_i,\boldsymbol{e}_j;-\sigma)TE(\boldsymbol{e}_i,\boldsymbol{e}_j;\sigma)\boldsymbol{e}_j=0.$$

并由此证明：对于任一矩阵 A，存在矩阵 $X=UF$，其中 U 为初等 Hermite 阵之积，F 为形如 $E(\boldsymbol{e}_i,\boldsymbol{e}_j;\sigma)$ 的初等矩阵之积，使得

$$X^{-1}AX=\mathrm{diag}(R_1,R_2,\cdots),$$

其中每个 $R_i(i=1,2,\cdots)$ 皆为对角线元素相等的上三角阵.

§3 矩阵分解

在§1中，我们已经引述了矩阵的 Jordan 分解(定理 1.1)和 Schur 分解(定理 1.2)，以及正规阵的酉对角分解(推论 1.1). 在这一节里，将论证矩阵的满秩分解，正定阵的三角分解，矩阵的奇异值分解和酉阵的对角酉块阵分解，它们是以后的讨论中要用到的.

定理 3.1(满秩分解). 设 $A\in\mathbf{C}_r^{m\times n}$，$r>0$. 则存在 $F\in\mathbf{C}_r^{m\times r}$ 和 $G\in\mathbf{C}_r^{r\times n}$，使得

$$A=FG. \tag{3.1}$$

证明:

由 $A \in \mathbf{C}_r^{m\times n}$ 可知,存在 $F \in \mathbf{C}_r^{m\times r}$, 使得

$$R(F) = R(A),$$

即 F 的列向量构成 $R(A)$ 的一组基底,因而 $A = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ 的每一个列向量可以唯一地用 F 的列向量表出,即

$$\boldsymbol{a}_i = F\boldsymbol{g}_i,\ \boldsymbol{g}_i \in \mathbf{C}^r,\ i = 1, \cdots, n.$$

令 $G = (\boldsymbol{g}_1, \cdots, \boldsymbol{g}_n)$, 则得到分解式(3.1);并且由

$$r \geqslant \operatorname{rank} G \geqslant \operatorname{rank}(FG) = r$$

可知 $G \in \mathbf{C}_r^{r\times n}$. □

注 3.1. 满秩分解(3.1)不是唯一的,因为对于任一 $Q \in \mathbf{C}_r^{r\times r}$, 若令 $F_1 = FQ$, $G_1 = Q^{-1}G$, 则显然有

$$A = F_1 G_1.$$

注 3.2. 利用初等下三角阵, 可以构造性地证明满秩分解(3.1)的存在性.证明如下:

设 $A^{(0)} \equiv A$ 的第 1 个非零列为第 j_1 列,第 j_1 列的第 1 个非零元素在第 i_1 行,则 $I_{1,i_1} A^{(0)}$ 的第 $(1, j_1)$ 元素不为零. 于是左乘以适当的单位下三角阵 L_1, 可使

$$A^{(1)} = L_1 I_{1,i_1} A^{(0)}$$

的第 j_1 列第 1 个元素以下全部为零,同时 $A^{(1)}$ 的前 $j_1 - 1$ 列亦全部保持为零. 然后找出第 1 个在第 1 行以下有非零元素的列, 假设它是第 j_2 列,并且第 (i_2, j_2) 元素是第 j_2 列中第 1 行以下第 1 个非零元素,则 $I_{2,i_2} A^{(1)}$ 的第 $(2, j_2)$ 元素不为零. 于是左乘以适当的单位下三角阵 L_2, 可使

$$A^{(2)} = L_2 I_{2,i_2} A^{(1)}$$

的第 j_2 列第 2 个元素以下全部为零, 并且 $A^{(2)}$ 的前 $j_2 - 1$列保持与 $A^{(1)}$ 相同. 依次类推,直到某一个 $A^{(r)}$, 它的第 r 行以下各行全部为零为止,这时

$$A^{(r)} = L_r I_{r,i_r} L_{r-1} I_{r-1,i_{r-1}} \cdots L_1 I_{1,i_1} A^{(0)}.$$

记

$$A^{(r)} = \begin{pmatrix} G \\ 0 \end{pmatrix}\begin{matrix} r \\ m-r \end{matrix}, \quad (L_r I_{r,i_r} \cdots L_1 I_{1,i_1})^{-1} = (\underset{r}{F}, \underset{n-r}{F'}),$$

则

$$A = FG, F \in \mathbb{C}_r^{m\times r}, G \in \mathbb{C}_r^{r\times n}. \quad \square$$

推论 3.1. 设 $A \in \mathbb{C}_r^{m\times n}$, $r > 0$. 则存在 $P \in \mathbb{C}_m^{m\times m}$ 和 $Q \in \mathbb{C}_n^{n\times n}$, 使得

$$A = P\begin{pmatrix} I^{(r)} & 0 \\ 0 & 0 \end{pmatrix} Q.$$

证明:

根据定理 3.1, 取 F' 与 G', 使得 $P = (F, F') \in \mathbb{C}_m^{m\times m}$ 和

$$Q = \begin{pmatrix} G \\ G' \end{pmatrix} \in \mathbb{C}_n^{n\times n}$$

即可。$\square$

定理 3.2. (正定阵的三角分解). 设 A 为正定阵. 则存在唯一的具有正对角线元素的下三角阵 L, 使得

$$A = LL^H.$$

证明:

对 A 的阶数施行数学归纳法. 当 A 为 1 阶时, 结论显然成立. 假定矩阵的阶数为 $n-1$ 时, 结论已成立, 现考虑任一 n 阶正定阵 A':

$$A' = \begin{pmatrix} A & \boldsymbol{a} \\ \boldsymbol{a}^H & \alpha \end{pmatrix}, \quad A \in \mathbb{C}^{(n-1)\times(n-1)}.$$

据 §1 关于正定阵的性质 3), $A > 0$. 要寻找的是具有正对角线元素的下三角阵 L', 满足 $L'L'^H = A'$. 记

$$L' = \begin{pmatrix} L & 0 \\ \boldsymbol{l}^H & \lambda \end{pmatrix}.$$

L' 应满足

$$LL^H = A, L\boldsymbol{l} = \boldsymbol{a}, \boldsymbol{l}^H\boldsymbol{l} + \lambda^2 = \alpha.$$

按照归纳法假设, 存在唯一的具有正对角线元素的下三角阵 L, 使得 $LL^H = A$; 随之可定出 $\boldsymbol{l} = L^{-1}\boldsymbol{a}$; 最后定出唯一的

$$\lambda = \sqrt{\alpha - \boldsymbol{l}^H \boldsymbol{l}}.$$

剩下只需证明 $\alpha - \boldsymbol{l}^H \boldsymbol{l} > 0$. 取

$$Q = \begin{pmatrix} I & 0 \\ -\boldsymbol{a}^H A^{-1} & 1 \end{pmatrix},$$

算出

$$QA'Q^H = \begin{pmatrix} A & 0 \\ 0 & \alpha - \boldsymbol{a}^H A^{-1} \boldsymbol{a} \end{pmatrix}.$$

据 § 1 关于正定阵的性质 5)和 3),有 $\alpha - \boldsymbol{a}^H A^{-1} \boldsymbol{a} > 0$, 即

$$\alpha - \boldsymbol{l}^H \boldsymbol{l} > 0. \quad \square$$

在讨论奇异值分解之前,首先引述

定义 3.1. 设 $A \in \mathbf{C}^{m \times n}$. $A^H A$ 的 n 个特征值的非负平方根叫做 A 的奇异值.

如果把 $A^H A$ 的特征值记作 $\lambda_i(A^H A), i = 1, \cdots, n$, 则 A 的奇异值就是指 $\sigma_i = \sqrt{\lambda_i(A^H A)},\ i = 1, \cdots, n$. A 的奇异值的全体,记作 $\sigma(A)$.

定理 3.3 (奇异值分解). 设 $A \in \mathbf{C}_r^{m \times n}$. 则存在酉阵 U 与 V, 使得

$$U^H A V = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{matrix} r \\ m - r \end{matrix}, \qquad (3.2)$$
$$\qquad\qquad\qquad r \quad n - r$$

其中 $\Sigma_1 = \text{diag}(\sigma_1, \cdots, \sigma_r), \sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$.

证明:

$A^H A$ 显然是半正定阵, 它的特征值皆为非负数, 记之为 $\sigma_1^2, \cdots, \sigma_n^2$, 并无妨假设它们满足 $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$ 和 $\sigma_{r+1} = \cdots = \sigma_n = 0$. 设 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n$ 分别是 $A^H A$ 的属于 $\sigma_1^2, \cdots, \sigma_n^2$ 的标准正交特征向量,并令 $V_1 = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_r)$, $V_2 = (\boldsymbol{v}_{r+1}, \cdots, \boldsymbol{v}_n)$, $V = (V_1, V_2)$和 $\Sigma_1 = \text{diag}(\sigma_1, \cdots, \sigma_r)$, 则有 $A^H A V_1 = V_1 \Sigma_1^2$,因而

$$V_1^H A^H A V_1 = \Sigma_1^2. \qquad (3.3)$$

再者,由 $A^H A V_2 = 0$ 知 $V_2^H A^H A V_2 = 0$, 因而

$$A V_2 = 0. \qquad (3.4)$$

令

$$U_1 = AV_1\Sigma_1^{-1}, \tag{3.5}$$

根据(3.3)，有 $U_1^H U_1 = I$；取 $U_2 \in \mathbb{C}^{m\times(m-r)}$，使 $U = (U_1, U_2)$ 为一酉阵. 则利用(3.3)—(3.5)，得到

$$U^H AV = \begin{pmatrix} U_1^H AV_1 & U_1^H AV_2 \\ U_2^H AV_1 & U_2^H AV_2 \end{pmatrix} = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}. \quad \square$$

注 3.3. 在 $\sigma_1 \geq \cdots \geq \sigma_r > 0$ 的限制下，分解式 (3.2) 中的 Σ_1 显然是唯一的. 但 U 与 V 并不唯一，因为由(3.2)可知 $U_1^H AV_1 = \Sigma_1$，于是可以分别在 $R(U_2)$ 与 $R(V_2)$ 中任取标准正交基构成 $\widetilde{U}_2$ 与 $\widetilde{V}_2$，代替 U_2 与 V_2；如果 $\sigma_1, \cdots, \sigma_r$ 中有 $\sigma_{i_1} = \cdots = \sigma_{i_p} (p > 1)$，则可在 $\boldsymbol{v}_{i_1}, \cdots, \boldsymbol{v}_{i_p}$ 所张成的子空间中，任取一组标准正交基 $\tilde{\boldsymbol{v}}_{i_1}, \cdots, \tilde{\boldsymbol{v}}_{i_p}$，代替 $\boldsymbol{v}_{i_1}, \cdots, \boldsymbol{v}_{i_p}$ (这时的 U_1 亦作相应的改变)；此外，任取一对对角酉阵 $D_1 = \mathrm{diag}(e^{i\theta_1}, \cdots, e^{i\theta_m})$ 和 $D_2 = \mathrm{diag}(e^{i\theta_1}, \cdots, e^{i\theta_n})$，令 $\widetilde{U} = UD_1$ 和 $\widetilde{V} = VD_2$，则亦有

$$\widetilde{U}^H A\widetilde{V} = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}.$$

下面的一条关于酉阵的对角酉块阵分解定理，对于研究子空间、正交投影以及广义逆的扰动规律很有用，它可以借助于奇异值分解加以证明.

定理 3.4. 设酉阵 $W \in \mathbb{C}^{n\times n}$ 分块为

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}\begin{matrix} l \\ n-l \end{matrix}, \quad 2l \leq n,$$
$$\qquad\begin{matrix} l & n-l \end{matrix}$$

则存在酉阵 $U = \mathrm{diag}(\underset{l}{U_{11}}, \underset{n-l}{U_{22}})$ 和 $V = \mathrm{diag}(\underset{l}{V_{11}}, \underset{n-l}{V_{22}})$，使得

$$U^H WV = \left(\begin{array}{c|cc} \Gamma & -\Sigma & 0 \\ \hline \Sigma & \Gamma & 0 \\ 0 & 0 & I \end{array}\right)\begin{matrix} l \\ l \\ n-2l \end{matrix}, \tag{3.6}$$
$$\qquad\begin{matrix} l & l & n-2l \end{matrix}$$

其中

$$\Gamma = \mathrm{diag}(\gamma_1, \cdots, \gamma_l) \geqslant 0, \Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_l) \geqslant 0.$$

证明:

分下列 4 步进行.

1）首先对 W_{11} 进行奇异值分解. 设

$$U_{11}^H W_{11} V_{11} = \Gamma$$

是 W_{11} 的奇异值分解,其中酉阵 $U_{11}, V_{11} \in \mathbf{C}^{l\times l}$,以及

$$\Gamma = \mathrm{diag}(\Gamma_1, I^{(l-k)}),\ \Gamma_1 = \mathrm{diag}(\gamma_1, \cdots, \gamma_k),$$

$0 \leqslant \gamma_1 \leqslant \cdots \leqslant \gamma_k < \gamma_{k+1} = \cdots = \gamma_l = 1,\ k \leqslant l$. 显然,矩阵 $\begin{pmatrix} W_{11} \\ W_{21} \end{pmatrix} V_{11}$ 有标准正交列,因此

$$I = \left(\begin{pmatrix} W_{11} \\ W_{21} \end{pmatrix} V_{11}\right)^H \left(\begin{pmatrix} W_{11} \\ W_{21} \end{pmatrix} V_{11}\right) = \Gamma^2 + (W_{21}V_{11})^H (W_{21}V_{11}).$$

即

$$(W_{21}V_{11})^H (W_{21}V_{11}) = \mathrm{diag}(I - \Gamma_1^2, 0^{(l-k)}).$$

可见 $W_{21}V_{11}$ 的前 k 列互相正交,后 $l-k$ 列为零.于是,将 $W_{21}V_{11}$ 的前 k 列标准化、并在右边补足与之标准正交的 $n-l-k$ 列之后可得一酉阵 $\hat{U}_{22} \in \mathbf{C}^{(n-l)\times(n-l)}$, 使得

$$\hat{U}_{22}^H W_{21} V_{11} = \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} \begin{matrix} l \\ n-2l \end{matrix},$$

其中

$$\left.\begin{aligned} &\Sigma = \mathrm{diag}(\underset{k}{\Sigma_1}, \underset{l-k}{0}),\ \Sigma_1 = \mathrm{diag}(\sigma_1, \cdots, \sigma_k) > 0, \\ &\qquad\qquad\qquad\qquad \sigma_{k+1} = \cdots = \sigma_l = 0. \end{aligned}\right\} \tag{3.7}$$

并且由

$$\mathrm{diag}(U_{11}, \hat{U}_{22})^H \begin{pmatrix} W_{11} \\ W_{21} \end{pmatrix} V_{11} = \begin{pmatrix} \Gamma \\ \Sigma \\ 0 \end{pmatrix}$$

的诸列皆为单位向量可知,必有

$$\gamma_i^2 + \sigma_i^2 = 1, i = 1, \cdots, l. \tag{3.8}$$

2）用类似于 1）中的办法, 可以确定出一个酉阵 $V_{22} \in$

$\mathbf{C}^{(n-l)\times(n-l)}$，使得

$$U_{11}^H W_{12} V_{22} = (T, 0),$$

其中 $T = \mathrm{diag}(\tau_1, \cdots, \tau_l), \tau_i \leqslant 0 (i = 1, \cdots, l)$. 并且由

$$U_{11}^H (W_{11}, W_{12}) \mathrm{diag}(V_{11}, V_{22}) = (\Gamma, T, 0)$$

的诸行皆为单位向量可知，必有 $\gamma_i^2 + \tau_i^2 = 1 (i = 1, \cdots, l)$，因而由(3.7)和(3.8)得出 $T = -\Sigma$.

3) 若令 $\hat{U} = \mathrm{diag}(U_{11}, \hat{U}_{22})$ 和 $V = \mathrm{diag}(V_{11}, V_{22})$，则上面已经证明了矩阵 $X = \hat{U}^H W V$ 有如下的分块：

$$X = \begin{pmatrix} \Gamma_1 & 0 & -\Sigma_1 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ \Sigma_1 & 0 & X_{33} & X_{34} & X_{35} \\ 0 & 0 & X_{43} & X_{44} & X_{45} \\ 0 & 0 & X_{53} & X_{54} & X_{55} \end{pmatrix} \begin{matrix} k \\ l-k \\ k \\ l-k \\ n-2l \end{matrix} . \tag{3.9}$$
$$\begin{matrix} k & l-k & k & l-k & n-2l \end{matrix}$$

注意到 X 是酉阵，它的诸列互相正交，利用 $\Sigma_1 > 0$ 便可推知(3.9)式右端的 $X_{33} = \Gamma_1$, X_{34}, X_{35}, X_{43} 与 X_{53} 皆为零矩阵. 因此

$$X = \begin{pmatrix} \Gamma_1 & 0 & -\Sigma_1 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ \Sigma_1 & 0 & \Gamma_1 & 0 & 0 \\ 0 & 0 & 0 & X_{44} & X_{45} \\ 0 & 0 & 0 & X_{54} & X_{55} \end{pmatrix},$$

其中 $\begin{pmatrix} X_{44} & X_{45} \\ X_{54} & X_{55} \end{pmatrix}$ 为酉阵.

4) 记

$$U_{33} = \begin{pmatrix} X_{44} & X_{45} \\ X_{54} & X_{55} \end{pmatrix} \in \mathbf{C}^{(n-l-k)\times(n-l-k)},$$

则有

$$\operatorname{diag}(I^{(l+k)}, U_{33}^H)X = \begin{pmatrix} \Gamma_1 & 0 & -\Sigma_1 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ \Sigma_1 & 0 & \Gamma_1 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix}$$

$$= \begin{pmatrix} \Gamma & -\Sigma & 0 \\ \Sigma & \Gamma & 0 \\ 0 & 0 & I \end{pmatrix} \begin{matrix} l \\ l \\ n-2l \end{matrix} .$$
$$\qquad l \qquad l \quad n-2l$$

再注意到

$$\operatorname{diag}(I^{(l+k)}, U_{33}^H)X = \operatorname{diag}(I^{(l+k)}, U_{33}^H)\hat{U}^H W V,$$

所以，若记

$$\begin{aligned} U &= \hat{U}\operatorname{diag}(I^{(l+k)}, U_{33}) \\ &= \operatorname{diag}(U_{11}, \hat{U}_{22})\operatorname{diag}(I^{(l)}, \operatorname{diag}(I^{(k)}, U_{33})) \\ &= \operatorname{diag}(U_{11}, \hat{U}_{22}\operatorname{diag}(I^{(k)}, U_{33})) \\ &= \operatorname{diag}(U_{11}, U_{22}), \end{aligned}$$

则 $U^H W V$ 具有(3.6)式所示的形式，其中 U 与 V 是定理中所述的对角块酉阵. □

利用定理 3.4，可得下述定理.

定理 3.5. 设 $X_1, Y_1 \in \mathbf{C}^{n\times l}$, $1 \leqslant l \leqslant n-1$, 且 $X_1^H X_1 = I$, $Y_1^H Y_1 = I$. 则必存在 $n\times n$ 酉阵 Q 和 $l\times l$ 酉阵 U_{11} 与 V_{11}, 使得

(1) 当 $2l \leqslant n$ 时，

$$QX_1U_{11} = \begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix} \begin{matrix} l \\ l \\ n-2l \end{matrix}, \quad QY_1V_{11} = \begin{pmatrix} \Gamma \\ \Sigma \\ 0 \end{pmatrix} \begin{matrix} l \\ l \\ n-2l \end{matrix}, \tag{3.10}$$

其中

$$\Gamma = \operatorname{diag}(\gamma_1, \cdots, \gamma_l), \Sigma = \operatorname{diag}(\sigma_1, \cdots, \sigma_l) \tag{3.11}$$

满足

$$\left.\begin{array}{r}0 \leqslant \gamma_1 \leqslant \cdots \leqslant \gamma_l, \sigma_1 \geqslant \cdots \geqslant \sigma_l \geqslant 0, \\ \gamma_i^2 + \sigma_i^2 = 1, i = 1, \cdots, l.\end{array}\right\} \quad (3.12)$$

(2) 当 $2l \geqslant n$ 时，

$$QX_1U_{11} = \begin{pmatrix} I & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix}\begin{matrix} n-l \\ 2l-n \\ n-l \end{matrix}, \quad QY_1V_{11} = \begin{pmatrix} \Gamma & 0 \\ 0 & I \\ \Sigma & 0 \end{pmatrix}\begin{matrix} n-l \\ 2l-n \\ n-l \end{matrix}$$
$$\begin{matrix} n-l & 2l-n \end{matrix} \qquad\qquad \begin{matrix} n-l & 2l-n \end{matrix}$$
$$(3.13)$$

其中

$$\Gamma = \operatorname{diag}(\gamma_1, \cdots, \gamma_{n-l}), \Sigma = \operatorname{diag}(\sigma_1, \cdots, \sigma_{n-l}) \quad (3.14)$$

满足

$$\left.\begin{array}{r}0 \leqslant \gamma_1 \leqslant \cdots \leqslant \gamma_{n-l}, \sigma_1 \geqslant \cdots \geqslant \sigma_{n-l} \geqslant 0, \\ \gamma_i^2 + \sigma_i^2 = 1, i = 1, \cdots, n-l.\end{array}\right\} \quad (3.15)$$

证明:

当 $2l \leqslant n$ 时，取 X_2，$Y_2 \in \mathbb{C}^{n \times (n-l)}$，使得 $X = (X_1, X_2)$ 与 $Y = (Y_1, Y_2)$ 为酉阵。令

$$W = X^H Y = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}\begin{matrix} l \\ n-l \end{matrix}.$$
$$\begin{matrix} l & n-l \end{matrix}$$

据定理 3.4，存在酉阵 U_{11}，U_{22}，V_{11} 与 V_{22}，使得(3.6)式成立。因此，若记

$$\hat{X}_i = X_i U_{ii} \ (i = 1, 2), \hat{X} = (\hat{X}_1, \hat{X}_2)$$

和

$$\hat{Y}_i = Y_i V_{ii} \ (i = 1, 2), \hat{Y} = (\hat{Y}_1, \hat{Y}_2),$$

则有

$$(\hat{X}_1, \hat{X}_2)^H (\hat{Y}_1, \hat{Y}_2) = \left(\begin{array}{c|cc} \Gamma & -\Sigma & 0 \\ \hline \Sigma & \Gamma & 0 \\ 0 & 0 & I \end{array}\right)\begin{matrix} l \\ l \\ n-2l \end{matrix}. \quad (3.16)$$
$$\begin{matrix} l & l & n-2l \end{matrix}$$

并且适当选取 U_{ii} 与 V_{ii}，可使(3.12)成立.

取 $Q=(\hat{X}_1,\hat{X}_2)^H$，显然 Q 是一酉阵. 于是得到

$$QX_1U_{11}=(\hat{X}_1,\hat{X}_2)^H\hat{X}_1=\begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix},$$

$$QY_1V_{11}=(\hat{X}_1,\hat{X}_2)^H\hat{Y}_1=\begin{pmatrix} \Gamma \\ \Sigma \\ 0 \end{pmatrix}.$$

这正是所要证明的(3.10)式.

同时有

$$\left.\begin{aligned} QX_2U_{22}=(\hat{X}_1,\hat{X}_2)^H\hat{X}_2&=\begin{pmatrix} 0 & 0 \\ I & 0 \\ 0 & I \end{pmatrix}\begin{matrix} l \\ l \\ n-2l \end{matrix}, \\ &\quad\ \ \begin{matrix} l & n-2l \end{matrix} \\ QY_2V_{22}=(\hat{X}_1,\hat{X}_2)^H\hat{Y}_2&=\begin{pmatrix} -\Sigma & 0 \\ \Gamma & 0 \\ 0 & I \end{pmatrix}\begin{matrix} l \\ l \\ n-2l \end{matrix}. \\ &\quad\ \ \begin{matrix} l & n-2l \end{matrix} \end{aligned}\right\} \tag{3.17}$$

注意到 X_2，$Y_2\in\mathbf{C}^{n\times(n-l)}$，$X_2^HX_2=Y_2^HY_2=I$，而 $2l\leqslant n$ 等价于 $2(n-l)\geqslant n$. 因此，如果把 l 换成 $n-l$，则(3.17)式给出了当 $2l\geqslant n$ 时，QX_1U_{11} 与 QY_1V_{11} 的标准形:

$$\begin{matrix}\begin{pmatrix} 0 & 0 \\ I & 0 \\ 0 & I \end{pmatrix}\begin{matrix} n-l \\ n-l \\ 2l-n \end{matrix}, \\ \begin{matrix} n-l & 2l-n \end{matrix}\end{matrix} \qquad \begin{matrix}\begin{pmatrix} -\Sigma & 0 \\ \Gamma & 0 \\ 0 & I \end{pmatrix}\begin{matrix} n-l \\ n-l \\ 2l-n \end{matrix}. \\ \begin{matrix} l & n-2l \end{matrix}\end{matrix}$$

再将它们的前 $n-l$ 行变号,并且左乘以一个适当的排列方阵,即可得到 (3.13). □

注 3.4. 定理 3.5 的几何意义. 设 $\mathscr{X}_1$ 与 $\mathscr{Y}_1$ 是 $\mathbf{C}^n$ 中的两个 l 维列空间,则可以通过一个适当的酉交换 Q,使得

(1) 当 $2l\leqslant n$ 时，(3.10)—(3.12)所示

$$\begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix} \text{和} \begin{pmatrix} \Gamma \\ \Sigma \\ 0 \end{pmatrix}$$

的列向量，分别构成 $Q\mathscr{X}_1$ 和 $Q\mathscr{Y}_1$ 的标准正交基；

(2) 当 $2l\geqslant n$ 时，(3.13)—(3.15)所示的

$$\begin{pmatrix} I & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix} \text{和} \begin{pmatrix} \Gamma & 0 \\ 0 & I \\ \Sigma & 0 \end{pmatrix}$$

的列向量，分别构成 $Q\mathscr{X}_1$ 和 $Q\mathscr{Y}_1$ 的标准正交基。

习题

1. 设 $A\in\mathbf{C}_p^{m\times n}$. 试求出 $A_k\in\mathbf{C}_k^{m\times n}$ $(1\leqslant k\leqslant p)$，使得 $\|A-A_k\|_F=\min$ (注: $\|A\|_F\equiv\sqrt{\operatorname{tr}(A^HA)}$).

2. 证明: 任一方阵有极分解式 $A=HQ$，其中 Q 为酉阵，H 为半正定阵. 此外，若 A 为非奇异阵，则 H 为正定阵，并且上述分解是唯一的.

3. 设 $A\in\mathbf{C}^{n\times m}$, $B\in\mathbf{C}^{m\times l}$. 试证

$$\operatorname{rank}(AB)\geqslant\operatorname{rank}(A)+\operatorname{rank}(B)-m.$$

4. 设 $A\in{}^{m\times n}$, $B\in\mathbf{C}^{n\times p}$, $C\in\mathbf{C}^{p\times q}$. 试证

$$\operatorname{rank}(AB)+\operatorname{rank}(BC)\leqslant\operatorname{rank}(B)+\operatorname{rank}(ABC).$$

5. 试对定理 3.5 给出一个直接的证明(不利用定理 3.4).

§4 值　　域

定义 4.1. 设 $A\in\mathbf{C}^{n\times n}$. $\mathbf{C}$ 中的有界闭集

$$F(A)=\{\boldsymbol{x}^HA\boldsymbol{x}:\boldsymbol{x}\in\mathbf{C}^n,\|\boldsymbol{x}\|_2=1\}$$

叫做 A 的值域.

$F(A)$ 是有界闭集的证明，作为一个练习(本节习题 1).

从定义 4.1 出发，容易直接验证 $F(A)$ 的下列性质:

$$F(\alpha A+\beta I)=\alpha F(A)+\beta,\alpha,\beta\in\mathbf{C}; \tag{4.1}$$

$$\lambda(A)\subseteq F(A); \tag{4.2}$$

$$F(U^HAU)=F(A),U\text{ 为酉阵}; \tag{4.3}$$

$$F(A+B)\subseteq F(A)+F(B)\equiv\{\alpha+\beta:\alpha\in F(A),\beta\in F(B)\}. \tag{4.4}$$

Toeplitz (1918) 和 Hausdorff (1919) 最早研究 $F(A)$ 的几何性质. 下面的定理4.1是一个著名的结果，通常称之为Toeplitz-Hausdorff 定理. 定理 4.1 有许多不同的证法，下面的证明是作者给出的.

定理 4.1. 值域 $F(A)$ 是 C 中的一个凸集*.

证明:

令 $v(\boldsymbol{x})=\boldsymbol{x}^HA\boldsymbol{x}$，此处 $\boldsymbol{x}^H\boldsymbol{x}=1$. 考虑 $F(A)$ 内任意二点 $v_1=v(\boldsymbol{x}_1)$ 和 $v_2=v(\boldsymbol{x}_2)$.

1) 注意到，C 中的 Euclid 变换并不改变 C 中任一凸集的凸性，所以，可首先作一 Euclid 变换

$$u=e^{-i\arg(v_2-v_1)}(v-v_1),$$

将 v_1 与 v_2 分别变为原点和正实轴上的一点. 这等于把 $F(A)$变为

$$F(e^{-i\arg(v_2-v_1)}(A-v_1I))=e^{-i\arg(v_2-v_1)}(F(A)-v_1).$$

因此，无妨假设 $v_1=0$, $v_2=r>0$. 于是问题归结为求证: 对于任一点 $t\in(0,r)$，必有 $\boldsymbol{x}_t\in\mathbf{C}^n$，且满足 $\|\boldsymbol{x}_t\|_2=1$，使得 $v(\boldsymbol{x}_t)=t$.

2) 令

$$\tilde{\boldsymbol{x}}_t=\sigma(t)\boldsymbol{x}_1+te^{i\theta}\boldsymbol{x}_2,\quad \boldsymbol{x}_t=\frac{\tilde{\boldsymbol{x}}_t}{\|\tilde{\boldsymbol{x}}_t\|_2}, \tag{4.5}$$

其中 θ 与 $\sigma(t)$ 为待定实数. 于是有

$$v(\boldsymbol{x}_t)=\boldsymbol{x}_t^HA\boldsymbol{x}_t=\frac{t^2r+t\sigma(t)(e^{i\theta}\boldsymbol{x}_1^HA\boldsymbol{x}_2+e^{-i\theta}\boldsymbol{x}_2^HA\boldsymbol{x}_1)}{\|\tilde{\boldsymbol{x}}_t\|_2^2},$$

---

* C 中的子集 $\mathscr{S}$ 称为凸集，如果对于 $\mathscr{S}$ 内任意二点 s_1 与 s_2，必有

$$\tau s_1+(1-\tau)s_2\in\mathscr{S},\forall\tau\in[0,1].$$

$$\|\tilde{\boldsymbol{x}}_t\|_2^2 = \sigma^2(t) + t^2 + 2t\sigma(t)\mathrm{Re}(e^{i\theta}\boldsymbol{x}_1^H\boldsymbol{x}_2).$$

为了使得 $v(\boldsymbol{x}_t) = t$，即

$$tr + \sigma(t)(e^{i\theta}\boldsymbol{x}_1^H A\boldsymbol{x}_2 + e^{-i\theta}\boldsymbol{x}_2^H A\boldsymbol{x}_1)$$
$$= \sigma^2(t) + t^2 + 2t\sigma(t)\mathrm{Re}(e^{i\theta}\boldsymbol{x}_1^H\boldsymbol{x}_2), \tag{4.6}$$

首先必须

$$\mathrm{Im}(e^{i\theta}\boldsymbol{x}_1^H A\boldsymbol{x}_2 + e^{-i\theta}\boldsymbol{x}_2^H A\boldsymbol{x}_1) = 0. \tag{4.7}$$

记 $\boldsymbol{x}_1^H A\boldsymbol{x}_2 = r_1 e^{i\alpha_1}$, $\boldsymbol{x}_2^H A\boldsymbol{x}_1 = r_2 e^{i\alpha_2}$,代入(4.7)，$\theta$ 应满足

$$0 = \mathrm{Im}(r_1 e^{i(\alpha_1+\theta)} + r_2 e^{i(\alpha_2-\theta)})$$
$$= r_1\sin(\alpha_1+\theta) + r_2\sin(\alpha_2-\theta),$$

由此可定出

$$\theta = \mathrm{tg}^{-1}\left(\frac{r_2\sin\alpha_2 + r_1\sin\alpha_1}{r_2\cos\alpha_2 - r_1\cos\alpha_1}\right). \tag{4.8}$$

再将(4.8)代入(4.6)，$\sigma(t)$ 应满足

$$\sigma^2(t) + [2t\mathrm{Re}(e^{i\theta}\boldsymbol{x}_1^H\boldsymbol{x}_2) - r_1\cos(\alpha_1+\theta) - r_2\cos(\alpha_2-\theta)]\sigma(t)$$
$$- t(r-t) = 0. \tag{4.9}$$

因为 $t(r-t) > 0$，所以方程(4.9)必有二实根(一正一负) $\sigma_1(t)$ 与 $\sigma_2(t)$。今任取其一，比如 $\sigma_1(t)$，则由 $\sigma_1(t)$ 和(4.8)式所示的 θ，通过(4.5)定出 $\boldsymbol{x}_t$，即为所求.□

定义 4.2. 设 $\lambda_1,\cdots,\lambda_k\in\mathbf{C}$. 如果 $\mathscr{H}(\lambda_1,\cdots,\lambda_k)$是 $\mathbf{C}$ 中包含 $\lambda_1,\cdots,\lambda_k$ 的凸集，并且 $\mathbf{C}$ 中凡包含 $\lambda_1,\cdots,\lambda_k$ 的凸集必包含 $\mathscr{H}(\lambda_1,\cdots,\lambda_k)$，则 $\mathscr{H}(\lambda_1,\cdots,\lambda_k)$ 叫做 $\mathbf{C}$ 中包含 $\lambda_1,\cdots,\lambda_k$ 的最小凸集，或者叫做 $\lambda_1,\cdots,\lambda_k$ 的凸包.

定理 4.2 (凸包表示定理). 设 $\lambda_1,\cdots,\lambda_k\in\mathbf{C}$. 则

$$\mathscr{H}(\lambda_1,\cdots,\lambda_k) = \left\{\sum_{i=1}^k\theta_i\lambda_i : \theta_i\geqslant 0, i=1,\cdots,k, \sum_{i=1}^k\theta_i = 1\right\}.$$

证明:

1) 首先证明：若 $\mathscr{S}$ 为 $\mathbf{C}$ 内一凸集，$\lambda_i\in\mathscr{S}$, $i=1,\cdots,k$，则 $\sum_{i=1}^k\theta_i\lambda_i\in\mathscr{S}$，只要 $\theta_i\geqslant 0 (i=1,\cdots,k)$并且 $\sum_{i=1}^k\theta_i = 1$.

对 λ_i 的个数 k 施行数学归纳法.当 $k=1$ 时，结论显然成立.

假定结论对于 k 已成立，现考虑 $k+1$ 的情形．记 $\theta=\sum_{i=1}^{k}\theta_i$．当 $\theta=0$ 和 $\theta=1$ 时，结论显然成立，所以只需考虑 $0<\theta<1$．

由归纳法假设知 $\sum_{i=1}^{k}\frac{\theta_i}{\theta}\lambda_i\in\mathscr{S}$，因此对于凸集 $\mathscr{S}$，

$$\theta\left(\sum_{i=1}^{k}\frac{\theta_i}{\theta}\lambda_i\right)+(1-\theta)\lambda_{k+1}\in\mathscr{S},$$

即

$$\sum_{i=1}^{k+1}\theta_i\lambda_i\in\mathscr{S}.$$

2）记

$$\tilde{\mathscr{H}}=\left\{\sum_{i=1}^{k}\theta_i\lambda_i:\theta_i\geqslant 0,\ i=1,\cdots,k,\ \sum_{i=1}^{k}\theta_i=1\right\}.$$

因为 $\mathscr{H}(\lambda_1,\cdots,\lambda_k)$ 是 $\mathbf{C}$ 中包含 $\lambda_1,\cdots,\lambda_k$ 的凸集，所以由 1)知 $\tilde{\mathscr{H}}\subseteq\mathscr{H}(\lambda_1,\cdots,\lambda_k)$．另一方面，容易直接验证 $\tilde{\mathscr{H}}$ 是 $\mathbf{C}$ 中包含 $\lambda_1,\cdots,\lambda_k$ 的凸集，因此由凸包的定义 4.2，立即得到 $\mathscr{H}(\lambda_1,\cdots,\lambda_k)\subseteq\tilde{\mathscr{H}}$．所以 $\tilde{\mathscr{H}}=\mathscr{H}(\lambda_1,\cdots,\lambda_k)$．□

下述定理 4.3 说明了正规阵的值域的几何特性．

定理 4.3. 设 $A\in\mathbf{C}^{n\times n}$ 是正规阵，$\lambda(A)=\{\lambda_i\}$．则

$$F(A)=\mathscr{H}(\lambda_1,\cdots,\lambda_n). \tag{4.10}$$

证明:

利用正规阵 A 的分解式 $A=U^H\Lambda U$，其中 U 为酉阵，$\Lambda=\operatorname{diag}(\lambda_1,\cdots,\lambda_n)$，可知

$$\begin{aligned}F(A)&=\{(U\boldsymbol{x})^H\Lambda(U\boldsymbol{x}):\boldsymbol{x}\in\mathbf{C}^n,\|\boldsymbol{x}\|_2=1\}\\&=\{\boldsymbol{y}^H\Lambda\boldsymbol{y}:\boldsymbol{y}\in\mathbf{C}^n,\|\boldsymbol{y}\|_2=1\}\\&=\left\{\sum_{i=1}^{n}\theta_i\lambda_i:\theta_i\geqslant 0,i=1,\cdots,n,\sum_{i=1}^{n}\theta_i=1\right\}.\end{aligned}$$

再由定理 4.2，立即得到(4.10)式．□

关于值域的研究已经相当深入，并且研究的范围也正在不断扩大. 这可参阅 M. Goldberg 的综合性论文[95]以及 Bauer 等人的论文[44]、[64]、[74]、[111]、[112]和[144].

习题

1. 设 $A\in \mathbf{C}^{n\times n}$. 证明$A$的值域 $F(A)$ 是 $\mathbf{C}$ 中的有界闭集.

2. 试证：矩阵A的值域 $F(A)$ 是实轴上的一个区间的必要与充分条件是A为 Hermite 阵.

§5 Kronecker 乘积

5.1 基本概念

定义 5.1. 设 $A=(\alpha_{ij})\in \mathbf{C}^{m\times n}, B\in \mathbf{C}^{p\times q}$. 则矩阵

$$A\otimes B\equiv\begin{pmatrix}\alpha_{11}B & \alpha_{12}B\cdots\alpha_{1n}B\\ \vdots & \vdots \qquad \vdots\\ \alpha_{m1}B & \alpha_{m2}B\cdots\alpha_{mn}B\end{pmatrix}\in \mathbf{C}^{mp\times nq}$$

叫做A和B的 Kronecker 乘积，或者叫做A和B的直乘积或者张量积.

由定义 5.1 知,如果 $C=(\gamma_{ij})\in \mathbf{C}^{n\times k}$, $D\in \mathbf{C}^{q\times r}$, 则 $C\otimes D\in \mathbf{C}^{nq\times kr}$.

定理 5.1. 设 A, B, C 与D如上所述. 则

$$(A\otimes B)(C\otimes D)=AC\otimes BD. \tag{5.1}$$

证明:

记

$$(A\otimes B)(C\otimes D)=\begin{pmatrix}\alpha_{11}B\cdots\alpha_{1n}B\\ \vdots \qquad \vdots\\ \alpha_{m1}B\cdots\alpha_{mn}B\end{pmatrix}\begin{pmatrix}\gamma_{11}D\cdots\gamma_{1k}D\\ \vdots \qquad \vdots\\ \gamma_{n1}D\cdots\gamma_{nk}D\end{pmatrix}$$

$$=\begin{pmatrix}P_{11}\cdots P_{1k}\\ \vdots\\ P_{m1}\cdots P_{mk}\end{pmatrix}, P_{st}\in \mathbf{C}^{p\times r}, 1\leqslant s\leqslant m, 1\leqslant t\leqslant k.$$

对任一对 s,t: $1\leqslant s\leqslant m, 1\leqslant t\leqslant k$, 显然有

$$P_{st}=\sum_{j=1}^{n}(\alpha_{sj}B)(\gamma_{jt}D)=\left(\sum_{j=1}^{n}\alpha_{sj}\gamma_{jt}\right)BD;$$

而 $\sum_{j=1}^{n}\alpha_{sj}\gamma_{jt}$ 恰好是 AC 的第 (s,t) 元素，因此(5.1)式成立. □

利用定义 5.1 和定理 5.1，容易验证下述定理成立.

定理 5.2. 设 A 与 B 如上所述. 则 $A\otimes B$ 有下列性质:

1) $(A\otimes B)^T=A^T\otimes B^T$, $\overline{(A\otimes B)}=\bar{A}\otimes\bar{B}$,

$$(A\otimes B)^H=A^H\otimes B^H;$$

2) 若 A 与 B 均为非奇异阵，则 $A\otimes B$ 亦为非奇异阵，并且有 $(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}$;

3) 若 $A\in\mathbf{C}^{m\times m}$, $B\in\mathbf{C}^{n\times n}$, 则

$$\det(A\otimes B)=(\det A)^n(\det B)^m;$$

4) $\operatorname{rank}(A\otimes B)=\operatorname{rank}A\cdot\operatorname{rank}B$.

5.2 应用举例：线性矩阵方程

考虑矩阵方程

$$AX-XB=C, \tag{5.2}$$

其中 $A\in\mathbf{C}^{m\times m}$, $B\in\mathbf{C}^{n\times n}$ 与 $C\in\mathbf{C}^{m\times n}$ 是已知矩阵，$X\in\mathbf{C}^{m\times n}$ 是未知矩阵.

记 $X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)$, $C=(\mathbf{c}_1,\cdots,\mathbf{c}_n)$, $\boldsymbol{x}_i$ 与 $\mathbf{c}_i\in\mathbf{C}^m$, $i=1,\cdots,n$. 令 $\boldsymbol{x}=(\boldsymbol{x}_1^T,\cdots,\boldsymbol{x}_n^T)^T$, 记作 $\boldsymbol{x}=\sum_{i=1}^{n}{}^{\cdot}\boldsymbol{x}_i$. 类似地，令 $\mathbf{c}=(\mathbf{c}_1^T,\cdots,\mathbf{c}_n^T)^T$, 记作 $\mathbf{c}=\sum_{i=1}^{n}{}^{\cdot}\mathbf{c}_i$. 以下证明方程(5.2)可以表示成

$$(I^{(n)}\otimes A-B^T\otimes I^{(m)})\boldsymbol{x}=\mathbf{c}. \tag{5.3}$$

证明:

令 $Y=AX=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n)$, $Z=XB=(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_n)$, 并令

$y = \sum_{i=1}^{n} {}^{\cdot} y_i$, $z = \sum_{i=1}^{n} {}^{\cdot} z_i$.　显然有

$$y = \sum_{i=1}^{n} {}^{\cdot} A x_i = (I^{(n)} \otimes A) x; \tag{5.4}$$

另一方面，如果记 B 的第 (s,t) 元素为 $(B)_{st}$，则由

$$z_t = (x_1, \cdots, x_n) \begin{pmatrix} (B)_{1t} \\ \vdots \\ (B)_{nt} \end{pmatrix} = \sum_{s=1}^{n} (B)_{st} x_s$$

$$= \sum_{s=1}^{n} (B^T)_{ts} x_s = \sum_{s=1}^{n} (B^T)_{ts} I^{(m)} x_s$$

可知

$$z = \sum_{t=1}^{n} {}^{\cdot} z_t = \sum_{t=1}^{n} {}^{\cdot} \sum_{s=1}^{n} (B^T)_{ts} I^{(m)} x_s$$

$$= (B^T \otimes I^{(m)}) x. \tag{5.5}$$

因此，将 Y，Z 与 C 按同样的方式分别表示成 $\mathbf{C}^{mn}$ 中的向量 y, z 与 c 之后，利用(5.4)和(5.5)，立即得知方程(5.2)可以表示成(5.3)的形式. □

定理 5.3. 设 $A \in \mathbf{C}^{m \times m}$，$B \in \mathbf{C}^{n \times n}$. 如果 $\lambda(A) = \{\lambda_i\}_{i=1}^{m}$，$\lambda(B) = \{\mu_j\}_{j=1}^{n}$，则

$$\lambda(I^{(n)} \otimes A - B^T \otimes I^{(m)}) = \{\lambda_i - \mu_j : i = 1, \cdots, m; \; j = 1, \cdots, n\}. \tag{5.6}$$

证明:

取 A 与 B^T 的 Schur 分解

$$A = U^H T_A U, \quad B^T = V^H T_B V,$$

其中 U 与 V 为酉阵，T_A 与 T_B 为上三角阵，它们的对角线元素分别是 A 与 B 的特征值. 于是由定理 5.1 和定理 5.2 可以得到 $I^{(n)} \otimes A - B^T \otimes I^{(m)}$ 的 Schur 分解:

$$I^{(n)} \otimes A - B^T \otimes I^{(m)}$$
$$= V^{-1} I^{(n)} V \otimes U^{-1} T_A U - V^{-1} T_B V \otimes U^{-1} I^{(m)} U$$

$$= (V^{-1}\otimes U^{-1})(I^{(n)}\otimes T_A)(V\otimes U)$$
$$- (V^{-1}\otimes U^{-1})(T_B\otimes I^{(m)})(V\otimes U)$$
$$= (V\otimes U)^{-1}(I^{(n)}\otimes T_A - T_B\otimes I^{(m)})(V\otimes U).$$

其中 $I^{(n)}\otimes T_A - T_B\otimes I^{(m)}$ 为上三角阵，它的对角线元素为 $\lambda_i - \mu_j, i = 1, \cdots, m, j = 1, \cdots, n$. 因此(5.6)得正. □

由(5.2)、(5.3)与(5.6)立即得出

推论 5.1. 矩阵方程 (5.2) 存在唯一解的必要与充分条件是 $\lambda(A)\cap\lambda(B) = \phi$($\phi$ 表示空集).

习题

1. 证明定理 5.2 中的结论 1)—4).

2 设 $A = \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix} \in \mathbf{C}^{n\times n}$, A_1 与 A_2 为方阵. 如果 $\lambda_1 \in \lambda(A_1)$, 但 $\lambda_1 \notin \lambda(A_2)$, 则在 A 与 A_1 中与 λ_1 相对应的初等因子是相同的.

§6 广 义 逆

6.1 基本概念

设 $A \in \mathbf{C}^{n\times n}$. 如果 A 是非奇异阵，则 A 存在唯一的逆矩阵 X, 满足

$$AX = XA = I. \tag{6.1}$$

A 的逆矩阵 X 记作 A^{-1}. 如果方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的系数矩阵 A 是非奇异阵，则该方程组存在唯一解 $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.

这是线性代数中熟知的结论.

但是如果方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的系数矩阵 A 是奇异阵，或者是长方阵，或者 $A\boldsymbol{x} = \boldsymbol{b}$ 是矛盾方程组，问这时应该如何将该方程组在某种意义下的解通过矩阵 A 的某种求逆加以表示呢? 这就有必要研究如何把矩阵求逆的概念、理论和方法推广到奇异方阵和一般长方阵的情形. 当然，这种推广后所得到的逆矩阵，应该具备普通逆矩阵的若干性质；而且当 A 是非奇异阵时，推广后的逆矩阵应

该与 A^{-1} 相吻合.

由(6.1)启发我们,任一矩阵 A 的逆矩阵 X,应该满足

$$AXA = A \tag{6.2$_1$}$$

和

$$XAX = X; \tag{6.2$_2$}$$

此外, AX 与 XA 应为 Hermite 阵,即

$$(AX)^H = AX \tag{6.2$_3$}$$

和

$$(XA)^H = XA. \tag{6.2$_4$}$$

利用(6.2$_1$)—(6.2$_4$)中的一个或几个条件，可以定义出 A 的不同类型的广义逆矩阵. 比如,由条件(6.2$_1$)可以定义出 A 的一类广义逆,记作 $A^{\{1\}}$; 由条件(6.2$_1$)和(6.2$_2$)可以定义出一类广义逆，记作 $A^{\{1,2\}}$; 等等. 并且这样定义出的广义逆，在方程组的求解中，确有相应的应用.

设 $A\in\mathbf{C}^{m\times n}$ 的奇异值分解为 $A = U\begin{pmatrix}\Sigma_1 & 0\\ 0 & 0\end{pmatrix}V^H$,其中 U 与 V 为酉阵，$\Sigma_1 = \mathrm{diag}(\sigma_1,\cdots,\sigma_r) > 0$. 容易验证

$$\left.\begin{aligned}
A^{\{1\}} &= V\begin{pmatrix}\Sigma_1^{-1} & K\\ L & M\end{pmatrix}U^H,\\
A^{\{1,2\}} &= V\begin{pmatrix}\Sigma_1^{-1} & K\\ L & 0\end{pmatrix}U^H,\ \text{其中 } K=0 \text{ 或 } L=0,\\
A^{\{1,2,3\}} &= V\begin{pmatrix}\Sigma_1^{-1} & 0\\ L & 0\end{pmatrix}U^H,\\
A^{\{1,2,4\}} &= V\begin{pmatrix}\Sigma_1^{-1} & K\\ 0 & 0\end{pmatrix}U^H
\end{aligned}\right\} \tag{6.3}$$

和

$$A^{\{1,2,3,4\}} = V\begin{pmatrix}\Sigma_1^{-1} & 0\\ 0 & 0\end{pmatrix}U^H. \tag{6.4}$$

广义逆的概念，在本世纪初就提出来了，开始时叫做“pseudoinverse”. 后来，E. H. Moore (1920) 和 R. Penrose (1955)着

重研究了由(6.2_1)—(6.2_4) 4 个方程所定义出的广义逆．由于这样得到的广义逆有着广泛的应用，所以下面仅就这类广义逆进行讨论。关于广义逆的比较全面的论述，可参看吴文达的综合性论文[27]和 Ben-Israel 与 Greville 的专著[55]．

定理 6.1. 设 $A \in \mathbf{C}^{m\times n}$．则方程组$(6.2_1)$—$(6.2_4)$存在唯一解 $X \in \mathbf{C}^{n\times m}$．

证明:

存在性已由(6.4)证明．以下证明唯一性．

首先指出，方程(6.2_1)—(6.2_4)的解 X 适合下列关系式:

$$\left.\begin{array}{l} X = XX^HA^H,\ X = A^HX^HX, \\ A = X^HA^HA,\ A = AA^HX^H. \end{array}\right\} \tag{6.5}$$

事实上，由(6.2_1)—(6.2_4)可知

$$\begin{aligned} X &= XAX = X(AX)^H = XX^HA^H, \\ X &= XAX = (XA)^HX = A^HX^HX, \\ A &= AXA = (AX)^HA = X^HA^HA, \\ A &= AXA = A(XA)^H = AA^HX^H. \end{aligned}$$

现证方程(6.2_1)—(6.2_4)的解 X 的唯一性．假设另有一解 Y，则 Y 亦应满足(6.5)中诸式，于是有

$$\begin{aligned} Y &= A^HY^HY = XAA^HY^HY = XAY = XX^HA^HAY \\ &= XX^HA^H = X. \ \square \end{aligned}$$

定义 6.1. 设 $A \in \mathbf{C}^{m\times n}$．方程 (6.2_1)—(6.2_4) 的解 X 叫做 A 的广义逆，通常称为 Moore-Penrose 广义逆，记作 $A^\dagger$．方程(6.2_1)—(6.2_4)叫做 $A^\dagger$ 的定义方程．

利用矩阵的满秩分解（定理 3.1），可以给出广义逆的显式表示．结论如下．

定理 6.2. 如果 $A \in \mathbf{C}_r^{m\times n} (r > 0)$ 的满秩分解为

$$A = FG, \tag{6.6}$$

则

$$A^\dagger = G^H(F^HAG^H)^{-1}F^H. \tag{6.7}$$

证明:

首先指出，F^HAG^H 为非奇异阵．事实上，由(6.6)可得

$$F^HAG^H=(F^HF)(GG^H),$$

其中 F^HF 与 GG^H 均为 $r\times r$ 满秩方阵． 因此 $F^HAG^H\in C_r^{r\times r}$.

所以，(6.7)也可以表示为

$$A^\dagger=G^H(GG^H)^{-1}(F^HF)^{-1}F^H. \tag{6.8}$$

直接验证可知，(6.8)所示的$A^\dagger$是方程(6.2_1)—(6.2_4)的解．□

由方程(6.2_1)—(6.2_4)还可看出，当 $A\in C_n^{n\times n}$ 时，$A^\dagger=A^{-1}$. 因此，广义逆 $A^\dagger$ 是满秩方阵的逆的一种推广.

6.2. 基本性质

从(6.2_1)—(6.2_4)关于 A 及 $A^\dagger$ 的对称性立即得出

1) $(A^\dagger)^\dagger=A$.

分别将(6.2_1)—(6.2_4)各式的两端取共轭转置，有

$$A^HA^{\dagger H}A^H=A^H,\qquad A^{\dagger H}A^HA^{\dagger H}=A^{\dagger H},$$

$$(A^{\dagger H}A^H)^H=A^{\dagger H}A^H,$$

$$(A^HA^{\dagger H})^H=A^HA^{\dagger H}.$$

因此，有

2) $(A^H)^\dagger=(A^\dagger)^H$.

同理可得

3) $(A^T)^\dagger=(A^\dagger)^T$.

利用(6.2_1)与(6.2_2)可证

4) $\mathrm{rank}(A)=\mathrm{rank}(A^\dagger)=\mathrm{rank}(A^\dagger A)$.

通过直接验证可得

5) $(AA^H)^\dagger=A^{H\dagger}A^\dagger$,

$(A^HA)^\dagger=A^\dagger A^{H\dagger}$.

再利用关系式(6.5)，得到

6) $(AA^H)^\dagger AA^H=AA^\dagger$,

$(A^HA)^\dagger A^HA=A^\dagger A$.

从(6.8)式可以推出

7) 若 $A\in C_n^{m\times n}$，则 $A^\dagger=(A^HA)^{-1}A^H$,

若 $A\in C_m^{m\times n}$，则 $A^\dagger=A^H(AA^H)^{-1}$.

特别地，非零列向量 $\boldsymbol{a}=(\alpha_1,\cdots,\alpha_m)^T$ 的广义逆为

$$\boldsymbol{a}^\dagger=\frac{1}{\sum_{i=1}^{m}|\alpha_i|^2}(\bar{\alpha}_1,\cdots,\bar{\alpha}_m);$$

非零行向量 $\boldsymbol{a}=(\alpha_1,\cdots,\alpha_n)$ 的广义逆为

$$\boldsymbol{a}^\dagger=\frac{1}{\sum_{i=1}^{n}|\alpha_i|^2}(\alpha_1,\cdots,\alpha_n)^H.$$

再联系到(6.8)式，得到

8）若 $A\in C_r^{m\times n}$ 有满秩分解(6.6)，则

$$A^\dagger=G^\dagger F^\dagger.$$

此外还有

9）若U与V为酉阵，则

$$(UAV)^\dagger=V^HA^\dagger U^H.$$

10）若 $A=\begin{pmatrix}D & 0\\ 0 & 0\end{pmatrix}\in C_r^{m\times n}$，$D=\mathrm{diag}(\delta_1,\cdots,\delta_r)$，$\delta_i\neq 0$，$i=1,\cdots,r$. 则

$$A^\dagger=\begin{pmatrix}D^{-1} & 0\\ 0 & 0\end{pmatrix}\in C_r^{n\times m}.$$

值得指出的是，A^{-1} 的许多性质，$A^\dagger$ 已不再具备. 比如当 A 为方阵时，一般地说:

1） $(AB)^\dagger\neq B^\dagger A^\dagger$.

例:

$$A=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix},\ B=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}.$$

这时

$$(AB)^\dagger=\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}^\dagger=\left(\begin{pmatrix}1\\ 0\end{pmatrix}(1,1)\right)^\dagger=\frac{1}{2}\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix},$$

而

$$B^{\dagger}A^{\dagger}=\begin{pmatrix}1 & -1\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.$$

2) $AA^{\dagger}\neq A^{\dagger}A$.

例:

$$A=\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}.$$

有

$$A^{\dagger}=\left(\begin{pmatrix}1\\ 1\end{pmatrix}(1,0)\right)^{\dagger}=\frac{1}{2}\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}$$

和

$$AA^{\dagger}=\frac{1}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix},$$

但

$$A^{\dagger}A=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.$$

3) $(A^k)^{\dagger}\neq(A^{\dagger})^k$.

例:

$$A=\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix},\quad k=2.$$

有

$$(A^2)^{\dagger}=\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}^{\dagger}=\frac{1}{2}\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix},$$

但

$$(A^{\dagger})^2=\left(\frac{1}{2}\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}\right)^2=\frac{1}{4}\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}.$$

4) A 与 $A^{\dagger}$ 的非零特征值,并不互为倒数.

例如,考虑 2) 中的 A,有 $\lambda(A)=\{1,0\}$,但

$$\lambda(A^{\dagger})=\left\{\frac{1}{2},\ 0\right\}.$$

习题

1. 证明 $(A\otimes B)^\dagger = A^\dagger\otimes B^\dagger$.

2. 证明下列结论:

 1) $A^\dagger = (A^HA)^\dagger A^H = A^H(AA^H)^\dagger$,

 2) 设 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 为列向量,则

$$(\boldsymbol{a}\boldsymbol{b}^H)^\dagger = (\boldsymbol{a}^H\boldsymbol{a})^\dagger(\boldsymbol{b}^H\boldsymbol{b})^\dagger\boldsymbol{b}\boldsymbol{a}^H.$$

3. 若 A 为方阵,则 $A^\dagger=A$ 的必要与充分条件是 A^2 为 Hermite 幂等阵,并且 $\mathrm{rank}(A^2)=\mathrm{rank}(A)$ (所谓方阵 A 为幂等阵,是指 $A^2=A$).

4. 设 A 为正规阵,则 $AA^\dagger = A^\dagger A$, 并且对任一自然数 k,有 $(A^k)^\dagger=(A^\dagger)^k$.

5. 设 $A\in\mathbf{C}^{n\times n}$, 则 $AA^\dagger=A^\dagger A$ 的必要与充分条件是 $N(A)=N(A^H)(N(A)=\{\boldsymbol{x}\in\mathbf{C}^n: A\boldsymbol{x}=0\})$.

§7 投　影

首先引述有关子空间的几个概念.

设 $A\in\mathbf{C}^{m\times n}$, 则称

$$R(A)=\{\boldsymbol{x}\in\mathbf{C}^m:\boldsymbol{x}=A\boldsymbol{y},\boldsymbol{y}\in\mathbf{C}^n\}$$

为 A 的列空间(或者称 A 的象域). 此外,称

$$N(A)=\{\boldsymbol{y}\in\mathbf{C}^n:A\boldsymbol{y}=0\}$$

为 A 的零空间.

设 L 与 M 是 $\mathbf{C}^n$ 的二子空间,即 $L,M\subseteq\mathbf{C}^n$. 如果

(i) 对于每个 $\boldsymbol{x}\in\mathbf{C}^n$, 存在 $\boldsymbol{x}_1\in L$ 和 $\boldsymbol{x}_2\in M$, 使得 $\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2$;

(ii) 如果 $\boldsymbol{x}\in L\cap M$, 则必有 $\boldsymbol{x}=0$;

则称 $\mathbf{C}^n$ 是二个子空间 L 与 M 的直接和,并记作

$$\mathbf{C}^n=L\oplus M.$$

如果 $\mathbf{C}^n=L\oplus M$, 则 L 与 M 叫做互补的子空间. 如果 $\boldsymbol{x}\perp$

$\boldsymbol{y}, \forall \boldsymbol{x} \in L, \forall \boldsymbol{y} \in M$，则称$M$是$L$的正交补子空间，$L$是$M$的正交补子空间，记作 $M = L^{\perp}$ 或 $L = M^{\perp}$.

易知，如果 $\mathbf{C}^n = L \oplus M$，，则任一向量 $\boldsymbol{x} \in \mathbf{C}^n$，必可唯一地表示成

$$\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}, \ \boldsymbol{y} \in L, \ \boldsymbol{z} \in M. \tag{7.1}$$

引理 7.1. 对于任一 $A \in \mathbf{C}^{m \times n}$，下列关系式成立:

$$N(A) = R(A^H)^{\perp}, \quad R(A) = N(A^H)^{\perp}.$$

证明:

设 $\boldsymbol{x} \in \mathbf{C}^n$，$\boldsymbol{y} \in \mathbf{C}^m$，则由

$$\boldsymbol{y}^H(A\boldsymbol{x}) = (A^H\boldsymbol{y})^H\boldsymbol{x}$$

可知，如果 $\boldsymbol{x} \in N(A)$，即 $A\boldsymbol{x} = 0$，则有

$$(A^H\boldsymbol{y})^H\boldsymbol{x} = 0, \quad \forall \boldsymbol{y} \in \mathbf{C}^m,$$

即 $\boldsymbol{x} \in R(A^H)^{\perp}$；因此 $N(A) \subseteq R(A^H)^{\perp}$.

另一方面，如果 $\boldsymbol{x} \in R(A^H)^{\perp}$，则有

$$\boldsymbol{y}^H(A\boldsymbol{x}) = (A^H\boldsymbol{y})^H\boldsymbol{x} = 0, \quad \forall \boldsymbol{y} \in \mathbf{C}^m,$$

因而 $A\boldsymbol{x} = 0$，即 $\boldsymbol{x} \in N(A)$. 所以又有 $R(A^H)^{\perp} \subseteq N(A)$.

所以 $N(A) = R(A^H)^{\perp}$.

对 A^H 应用上述结果，即得到 $R(A) = N(A^H)^{\perp}$. □

7.1 幂等阵与投影

定义 7.1. 如果 $E \in \mathbf{C}^{n \times n}$ 适合 $E^2 = E$，则称E为幂等阵.

定理 7.1. 设 $E \in \mathbf{C}^{n \times n}$ 为幂等阵，则

1) E^H 与 $I - E$ 亦为幂等阵，

2) E 的 Jordan 标准形为 $\begin{pmatrix} I^{(r)} & 0 \\ 0 & 0 \end{pmatrix}$，$r \leqslant n$，

3) $\operatorname{rank}(E) = \operatorname{tr}(E)$，

4) $R(I - E) = N(E)$，

5) $N(I - E) = R(E)$，

6) 如果E有满秩分解 $E = FG$，则 $GF = I$.

证明:

1)与 2)可从幂等阵的定义 7.1 直接导出，3)是 2) 的推论，5)是 1)与 4)的推论. 以下只证 4)与 6).

关于 4).

如果 $\boldsymbol{x} \in R(I-E)$，则存在 $\boldsymbol{y} \in \mathbf{C}^n$，使得 $\boldsymbol{x}=(I-E)\boldsymbol{y}$，从而 $E\boldsymbol{x}=E(I-E)\boldsymbol{y}=(E-E^2)\boldsymbol{y}=0$，即 $\boldsymbol{x} \in N(E)$. 所以有 $R(I-E) \subseteq N(E)$. 反之，如果 $\boldsymbol{x} \in N(E)$，则有 $E\boldsymbol{x}=0$，于是 $(I-E)\boldsymbol{x}=\boldsymbol{x}$，此示 $\boldsymbol{x} \in R(I-E)$，所以又有 $N(E) \subseteq R(I-E)$. 因此 $R(I-E)=N(E)$.

关于 6).

由 $E^2=E$ 知 $F(GF-I)G=0$，因而

$$F^H F(GF-I)GG^H=0.$$

注意到 $F^H F$ 与 GG^H 均为非奇异阵，所以有 $GF-I=0$，即 $GF=I$. □

定义 7.2. 设 $\mathbf{C}^n=L \oplus M$，$\boldsymbol{x} \in \mathbf{C}^n$ 有分解式

$$\boldsymbol{x}=\boldsymbol{y}+\boldsymbol{z},\ \boldsymbol{y} \in L,\ \boldsymbol{z} \in M.$$

则 $\boldsymbol{y}$ 叫做 $\boldsymbol{x}$ 沿 M 到 L 的投影. 如果用 $P_{L,M}$ 表示相应的由 $\mathbf{C}^n$ 到 L 上的映射，则 $P_{L,M}$ 叫做沿 M 到 L 上的投影变换，或投影算子.

由定义 7.2 可立即得到

推论 7.1. 若 $\mathbf{C}^n=L \oplus M$，则 $\mathbf{C}^n$ 内任一向量 $\boldsymbol{x}$ 的唯一分解式(7.1)中的 $\boldsymbol{y}$ 与 $\boldsymbol{z}$ 由

$$\boldsymbol{y}=P_{L,M}\boldsymbol{x},\ \boldsymbol{z}=(I-P_{L,M})\boldsymbol{x}$$

给出.

容易验证：$P_{L,M}$ 是一个线性变换，即

$$P_{L,M}(\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2)=\alpha_1 P_{L,M}\boldsymbol{x}_1+\alpha_2 P_{L,M}\boldsymbol{x}_2,$$

$$\forall \boldsymbol{x}_1,\ \boldsymbol{x}_2 \in \mathbf{C}^n,\ \forall \alpha_1,\ \alpha_2 \in \mathbf{C}.$$

与此线性变换 $P_{L,M}$ 相对应的矩阵，仍记作 $P_{L,M}$.

下述定理说明了幂等阵与投影算子的关系.

定理 7.2. 如果 $E \in \mathbf{C}^{n \times n}$ 是一个幂等阵，则

$$\mathbf{C}^n=R(E) \oplus N(E),$$

并且

$$E = P_{R(E),N(E)};$$

反之，如果 $\mathbf{C}^n = L \oplus M$，则存在唯一的幂等阵 $P_{L,M}$ 使得

$$R(P_{L,M}) = L,\ N(P_{L,M}) = M.$$

证明:

设 $E = \mathbf{C}^{n\times n}$ 是一幂等阵，则一方面，有

$$\boldsymbol{x} = E\boldsymbol{x} + (I - E)\boldsymbol{x},\ \forall \boldsymbol{x} \in \mathbf{C}^n,$$

其中 $E\boldsymbol{x} \in R(E)$，$(I - E)\boldsymbol{x} \in R(I - E) = N(E)$（见定理 7.1 4））。另一方面，如果 $\boldsymbol{z} \in R(E) \cap N(E)$，即存在 $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbf{C}^n$，使得

$$\boldsymbol{z} = E\boldsymbol{x}_1 = (I - E)\boldsymbol{x}_2,$$

则必有

$$\boldsymbol{z} = E\boldsymbol{x}_1 = E^2\boldsymbol{x}_1 = E(I - E)\boldsymbol{x}_2 = 0.$$

因此 $\mathbf{C}^n = R(E) \oplus N(E)$. 根据定义 7.2，$E\boldsymbol{x}$ 正是 $\boldsymbol{x}$ 沿 $N(E)$ 到 $R(E)$ 上的投影，所以

$$E = P_{R(E),N(E)}.$$

反之，如果 $\mathbf{C}^n = L \oplus M$，则可分别在 L 与 M 内任取基底 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_l\}$ 与 $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\}$，$l + m = n$. 于是，如果沿着 M 到 L 上的投影算子记作 $P_{L,M}$，则它应由

$$\begin{cases} P_{L,M}\boldsymbol{x}_i = \boldsymbol{x}_i, & i = 1, \cdots, l \\ P_{L,M}\boldsymbol{y}_j = 0, & j = 1, \cdots, m \end{cases} \tag{7.2}$$

确定. 令 $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_l)$，$Y = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_m)$，(7.2) 可表示成

$$P_{L,M}(X, Y) = (X, 0). \tag{7.3}$$

因为 $(X, Y) \in \mathbf{C}^{n\times n}$ 为非奇异阵，所以由(7.3)可解出(7.2)的唯一解

$$P_{L,M} = (X, O)(X, Y)^{-1}.$$

此外，由(7.2)可知

$$P_{L,M}(X, O) = (X, O),$$

从而有

$$\begin{aligned} P_{L,M}^2 &= P_{L,M}(X, O)(X, Y)^{-1} \\ &= (X, O)(X, Y)^{-1} = P_{L,M}, \end{aligned}$$

即 $P_{L,M}$ 为幂等阵. □

例 7.1. $E = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$是幂等阵. 有 $R(E) = R\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)$, $N(E) = R\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$. $\mathbf{R}^2$ 内任一向量 $\boldsymbol{x}$ 沿 $N(E)$ 到 $R(E)$ 上的投影

$$P_{R(E),N(E)}\boldsymbol{x} = E\boldsymbol{x}.$$

如图 1-2 所示.

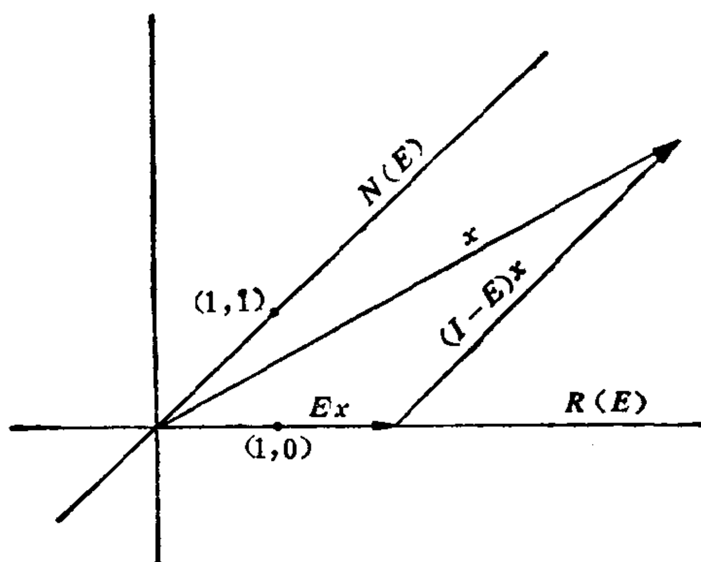


图 1-2

上面所说的投影是斜投影，它与幂等阵相对应. 而普通所说的投影是正交投影,见下段.

7.2 正交投影

定理 7.3. 设 L 是 $\mathbf{C}^n$ 的子空间， $\boldsymbol{x}$ 是 $\mathbf{C}^n$ 内任一向量. 则在 L 中存在唯一的向量 $\boldsymbol{u}_x$，使得

$$\|\boldsymbol{x} - \boldsymbol{u}_x\|_2 < \|\boldsymbol{x} - \boldsymbol{u}\|_2, \ \forall \boldsymbol{u} \in L, \boldsymbol{u} \neq \boldsymbol{u}_x. \tag{7.4}$$

证明:

令 $L^\perp$ 表示 L 的正交补子空间. 容易验证

$$\mathbf{C}^n = L \oplus L^\perp.$$

因此,对于任一 $\boldsymbol{x} \in \mathbf{C}^n$，有唯一分解(见图 1-3)

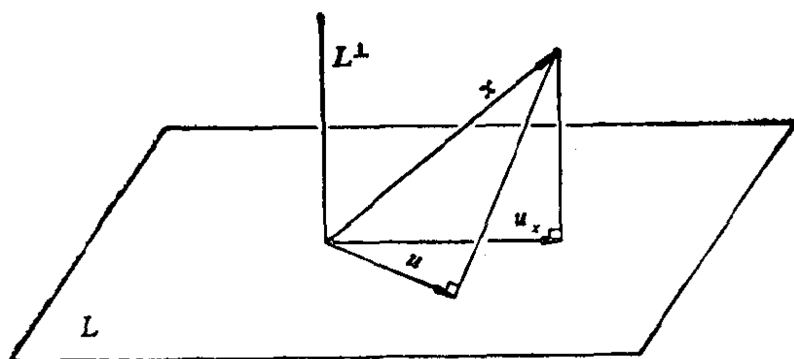


图 1-3

$$\boldsymbol{x}=\boldsymbol{u}_x+(\boldsymbol{x}-\boldsymbol{u}_x),\boldsymbol{u}_x\in L,\boldsymbol{x}-\boldsymbol{u}_x\in L^{\perp}. \tag{7.5}$$

于是，对任一 $\boldsymbol{u}\in L$，有

$$\begin{aligned}\|\boldsymbol{x}-\boldsymbol{u}\|_2^2&=\|(\boldsymbol{u}_x-\boldsymbol{u})+(\boldsymbol{x}-\boldsymbol{u}_x)\|_2^2\\&=\|\boldsymbol{u}_x-\boldsymbol{u}\|_2^2+\|\boldsymbol{x}-\boldsymbol{u}_x\|_2^2.\end{aligned}$$

显然，在 L 中存在唯一的 $\boldsymbol{u}=\boldsymbol{u}_x$，使 $\|\boldsymbol{x}-\boldsymbol{u}\|_2$ 达到极小值 $\|\boldsymbol{x}-\boldsymbol{u}_x\|_2$，并且不等式(7.4)成立. □

对任一 $\boldsymbol{x}\in \mathrm{C}^n$，由定理 7.3 在 L 中所唯一确定的 $\boldsymbol{u}_x$（即由分解式(7.5)所唯一确定的 $\boldsymbol{u}_x$），叫做 $\boldsymbol{x}$ 沿 $L^{\perp}$ 到 L 的正交投影. 如果用 P_L 表示相应的由 C^n 到 L 上的映射，则 P_L 叫做沿 $L^{\perp}$ 到 L 的正交投影变换，或正交投影算子.

同时，由(7.5)所唯一确定的 $\boldsymbol{x}-\boldsymbol{u}_x=(I-P_L)\boldsymbol{x}$ 叫做 $\boldsymbol{x}$ 沿 L 到 $L^{\perp}$ 的正交投影. 如果用 $P_{L^{\perp}}$ 表示相应的由 C^n 到 $L^{\perp}$ 上的映射，则

$$P_{L^{\perp}}=I-P_L$$

叫做沿 L 到 $L^{\perp}$ 的正交投影变换，或正交投影算子.

与线性变换 P_L 和 $P_{L^{\perp}}$ 相对应的矩阵，仍分别记作 P_L 和 $P_{L^{\perp}}$.

下述定理说明了正交投影算子与 Hermite 幂等阵的关系. 所谓矩阵 P 是 Hermite 幂等阵，就是指 P 既是幂等阵，又是 Hermite 阵.

定理 7.4. 设 $\mathrm{C}^n=L\oplus M$. 则 $P_{L,M}$ 是 Hermite 阵的必要

与充分条件是 $M = L^{\perp}$.

证明:

由引理 7.1 和定理 7.2 可知,

$$\left.\begin{aligned} R(P_{L,M}^{H}) &= N(P_{L,M})^{\perp} = M^{\perp}, \\ N(P_{L,M}^{H}) &= R(P_{L,M})^{\perp} = L^{\perp}. \end{aligned}\right\} \tag{7.6}$$

因为 $P_{L,M}^{H}$ 是幂等阵,据定理 7.2, $R(P_{L,M}^{H})$ 与 $N(P_{L,M}^{H})$ 互补,即 $M^{\perp}$ 与 $L^{\perp}$ 互补.

再据定理7.2,存在幂等阵 $P_{M^{\perp},L^{\perp}}$, 使得

$$R(P_{M^{\perp},L^{\perp}}) = M^{\perp},\ N(P_{M^{\perp},L^{\perp}}) = L^{\perp}, \tag{7.7}$$

并且这个幂等阵是唯一的. 比较(7.6) 与 (7.7) 可知(见本节习题3)

$$P_{L,M}^{H} = P_{M^{\perp},L^{\perp}}.$$

于是, $P_{L,M}$ 是 Hermite 阵的必要与充分条件是

$$P_{L,M} = P_{M^{\perp},L^{\perp}}.$$

再由子空间与投影算子的一一对应关系得出 $P_{L,M}$ 是 Hermite 阵的必要与充分条件是 $M = L^{\perp}$. □

定理 7.2 与定理 7.4 表明: 正交投影算子 P_L 必是 Hermite 幂等阵;反之,如果 P 是一个 Hermite 幂等阵,则必存在 $\mathbf{C}^n$ 的一个正交补分解: $\mathbf{C}^n = L \oplus L^{\perp}$, 使得 $P = P_L$.

这样就建立了正交投影算子与 Hermite 幂等阵之间的一一对应关系.

7.3 $AA^{\dagger}$ 与 $A^{\dagger}A$ 的几何意义

首先证明一条关于 $AA^{\dagger}$ 与 $A^{\dagger}A$ 的列空间和零空间的命题.

引理 7.2. 设 $A \in \mathbf{C}^{m \times n}$. 则有

$$\begin{aligned} &R(AA^{\dagger}) = R(AA^{H}) = R(A), \\ &R(A^{\dagger}A) = R(A^{H}A) = R(A^{\dagger}) = R(A^{H}) \end{aligned} \tag{7.8}$$

和

$$\left.\begin{aligned} N(AA^\dagger) = N(AA^H) = N(A^\dagger) = N(A^H), \\ N(A^\dagger A) = N(A^H A) = N(A). \end{aligned}\right\} \quad (7.9)$$

证明:

设 $A \in \mathbf{C}_r^{m\times n}$. 进行奇异值分解: $A = U\begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} V^H$,其中 $U = (\underset{r}{U_1}, \underset{m-r}{U_2})$ 和 $V = (\underset{r}{V_1}, \underset{n-r}{V_2})$ 是酉阵, $\Sigma_1 = \mathrm{diag}(\sigma_1, \cdots, \sigma_r) > 0$. 于是得到

$$A = U_1\Sigma_1 V_1^H, \ A^H = V_1\Sigma_1 U_1^H, \ A^\dagger = V_1\Sigma_1^{-1}U_1^H$$

和

$$AA^H = U_1\Sigma_1^2 U_1^H, \quad A^H A = V_1\Sigma_1^2 V_1^H,$$
$$AA^\dagger = U_1U_1^H, \ A^\dagger A = V_1V_1^H.$$

容易看出(见本节习题 5)

$R(A) = R(U_1), R(A^H) = R(V_1), R(A^\dagger) = R(V_1),$

$R(AA^H) = R(U_1), R(A^H A) = R(V_1), R(AA^\dagger) = R(U_1),$

$R(A^\dagger A) = R(V_1)$

和

$N(A) = R(V_2), N(A^H) = R(U_2), N(A^\dagger) = R(U_2),$

$N(AA^H) = R(U_2), N(A^H A) = R(V_2), N(AA^\dagger) = R(U_2),$

$N(A^\dagger A) = R(V_2).$

由此立即得到(7.8)和(7.9). □

现在考察 $AA^\dagger$ 与 $A^\dagger A$ 的几何意义(设 $A \in \mathbf{C}^{m\times n}$).

显然, $AA^\dagger \in \mathbf{C}^{m\times m}$ 是 Hermite 幂等阵. 据定理 7.2,有

$$\mathbf{C}^m = R(AA^\dagger) \oplus N(AA^\dagger),$$

并且 $AA^\dagger$ 是沿着 $N(AA^\dagger)$ 到 $R(AA^\dagger)$ 的投影算子. 由(7.8)和(7.9)知, $R(AA^\dagger) = R(A), N(AA^\dagger) = N(A^H)$. 再由引理 7.1 知, $N(A^H) = R(A)^\perp$. 所以, 根据定理 7.4, AA^+ 是到 $R(A)$ 上的正交投影算子,通常记为

$$P_{R(A)} = AA^\dagger, \text{ 或 } P_A = AA^\dagger.$$

再者, $A^\dagger A \in \mathbf{C}^{n\times n}$ 显然也是 Hermite 幂等阵. 据定理 7.2,

有

$$\mathbf{C}^n = R(A^\dagger A) \oplus N(A^\dagger A),$$

并且 $A^\dagger A$ 是沿着 $N(A^\dagger A)$ 到 $R(A^\dagger A)$ 的投影算子．由(7.8)和(7.9)知，$R(A^\dagger A) = R(A^H)$，$N(A^\dagger A) = N(A)$．再由引理 7.1 知，$N(A) = R(A^H)^\perp$． 所以，根据定理 7.4，$A^\dagger A$ 是到 $R(A^H)$ 的正交投影算子，通常记为

$$P_{R(A^H)} = A^\dagger A, \text{ 或 } P_{A^H} = A^\dagger A.$$

7.4 应用举例：线性最小二乘问题

设 $A \in \mathbf{C}^{m\times n}$，$\boldsymbol{b} \in \mathbf{C}^m$． 所谓线性最小二乘问题，是指求 $\boldsymbol{x} \in \mathbf{C}^n$，使得

$$\rho(\boldsymbol{x}) \equiv \|A\boldsymbol{x} - \boldsymbol{b}\|_2 = \min. \tag{7.10}$$

定理 7.5. 线性最小二乘问题(7.10)的解由

$$\boldsymbol{x} = A^\dagger b + (I^{(n)} - P_{A^H})\boldsymbol{z} \tag{7.11}$$

给出， 其中 $\boldsymbol{z}$ 表示 $\mathbf{C}^n$ 中任一向量； 当 A 的列向量线性无关时，(7.10)有唯一解 $\boldsymbol{x} = A^\dagger \boldsymbol{b}$． 此外，(7.10) 有唯一的最小范数解 $\boldsymbol{x} = A^\dagger \boldsymbol{b}$，它同时满足(7.10)和

$$\|\boldsymbol{x}\|_2 = \min.$$

证明:

首先注意到， $A\boldsymbol{x} \in R(A), \forall \boldsymbol{x} \in \mathbf{C}^n$． 因此，可以把向量 $\boldsymbol{b}$ 分解成 $\boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{b}_2$， 其中 $\boldsymbol{b}_1 = P_A \boldsymbol{b}$ 是 $\boldsymbol{b}$ 到 $R(A)$ 的正交投影，$\boldsymbol{b}_2 = \boldsymbol{b} - \boldsymbol{b}_1$． 由

$$\rho^2(\boldsymbol{x}) \equiv \|\boldsymbol{b} - A\boldsymbol{x}\|_2^2 = \|\boldsymbol{b}_1 - A\boldsymbol{x}\|_2^2 + \|\boldsymbol{b}_2\|_2^2$$

可知，使 $\rho(\boldsymbol{x}) = \min$ 的必要与充分条件是 $A\boldsymbol{x} = \boldsymbol{b}_1$，显然， 这样的 $\boldsymbol{x}$ 是存在的，因为 $\boldsymbol{b}_1 \in R(A)$.

于是(7.10)的解 $\boldsymbol{x}$ 必须而且只需满足

$$A\boldsymbol{x} = AA^\dagger \boldsymbol{b}.$$

现设

$$\boldsymbol{x} = A^\dagger \boldsymbol{b} + \boldsymbol{y}. \tag{7.12}$$

从
$$Ay = A(x - A^{\dagger}b) = Ax - P_A b = b_1 - b_1 = 0$$
可知 $y \in N(A)$，再由引理 7.1，$N(A) = R(A^H)^{\perp}$，所以 $y \in R(A^H)^{\perp}$. 而 $R(A^H)^{\perp}$ 中的向量 y 必可表示成
$$y = z - P_{A^H} z = (I^{(n)} - P_{A^H}) z, z \in \mathbf{C}^n, \tag{7.13}$$
将(7.13)代入(7.12)，便得到(7.10)的通解(7.11).

当 A 的列向量线性无关时，因有 $P_{A^H} = I$，所以 (7.10) 有唯一解 $x = A^{\dagger}b$.

此外，因为 $A^{\dagger}b$ 与 $(I - P_{A^H})z$ 正交，所以由
$$\|x\|_2^2 = \|A^{\dagger}b\|_2^2 + \|(I - P_{A^H})z\|_2^2$$
立即得出(7.10)的最小范数解为 $x = A^{\dagger}b$. □

习题

1. 证明 n 行 n 列幂等阵的 Jordan 标准形为 $\begin{pmatrix} I^{(r)} & 0 \\ 0 & 0 \end{pmatrix}$, $r \leqslant n$.

2. 如果 $\text{rank}(A) = \text{rank}(BA)$，则
$$\text{rank}(AC) = \text{rank}(BAC).$$

3. 设 A 与 B 为同阶幂等阵，并且满足 $R(A) = R(B)$ 和 $N(A) = N(B)$. 试证 $A = B$.

4. 试证：任一正交投影算子 P 必可表为 $P = VV^H$，其中 V 满足 $V^H V = I$.

5. 设 $A \in \mathbf{C}^{m \times n}$ 的奇异值分解为 $A = U \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} V^H$，其中 $U = (\underset{r}{U_1}, \underset{m-r}{U_2})$ 与 $V = (\underset{r}{V_1}, \underset{n-r}{V_2})$ 为酉阵，$\Sigma_1 = \text{diag}(\sigma_1, \cdots, \sigma_r)$，$\sigma_1, \cdots, \sigma_r > 0$. 试证：
$$R(A) = R(U_1),\ N(A) = R(V_2).$$

6. 设 A 是一奇异方阵，$\{u_1, \cdots, u_k\}$ 与 $\{x_1, \cdots, x_k\}$ 分别是

$N(A^H)$ 与 $N(A)$ 的标准正交基，$\alpha_1,\cdots,\alpha_k$ 是 k 个非零复数。试证：矩阵

$$A_0 = A + \sum_{i=1}^{k} \alpha_i \boldsymbol{u}_i \boldsymbol{x}_i^H$$

是非奇异阵，并且

$$A_0^{-1} = A^{\dagger} + \sum_{i=1}^{k} \frac{1}{\alpha_i} \boldsymbol{x}_i \boldsymbol{u}_i^H.$$

7. 试证：最小二乘问题

$$\begin{cases} \|AXB - C\|_F = \min \\ \|X\|_F = \min \end{cases}$$

的解为 $X = A^{\dagger} C B^{\dagger}$。

§8 行 列 式

本节讨论 Binet-Cauchy 公式和 Hadamard 不等式。

8.1 Binet-Cauchy 公式

设 $A = (\alpha_{ij}) \in \mathbb{C}^{m\times n}$。分别在 $1,\cdots,m$ 中和在 $1,\cdots,n$ 中取 p 个互不相同的数 $i_1,\cdots,i_p$ 和 $k_1,\cdots,k_p$。记

$$A\begin{pmatrix} i_1\cdots i_p \\ k_1\cdots k_p \end{pmatrix} = \det\begin{pmatrix} \alpha_{i_1k_1} & \cdots & \alpha_{i_1k_p} \\ \vdots & & \vdots \\ \alpha_{i_pk_1} & \cdots & \alpha_{i_pk_p} \end{pmatrix}, \quad A\begin{pmatrix} i_1\cdots i_p \\ k_1\cdots k_p \end{pmatrix}$$

叫做 A 的一个 p 阶子式。

定理 8.1 (Binet-Cauchy 公式)· 设 $A = (\alpha_{ij}) \in \mathbb{C}^{m\times n}$，$B = (\beta_{ij}) \in \mathbb{C}^{n\times m}$。则

$$\det AB = \begin{cases} 0 & \text{如果 } m > n, \\ \sum A\begin{pmatrix} 1\cdots m \\ k_1\cdots k_m \end{pmatrix} B\begin{pmatrix} k_1\cdots k_m \\ 1\cdots m \end{pmatrix} & \text{如果 } m \leqslant n. \end{cases} \tag{8.1}$$

上式中的和号 $\sum$ 取 $1 \leqslant k_1 \leqslant \cdots \leqslant k_m \leqslant n$。

证明：

如果 $m > n$，则由 $AB \in \mathbb{C}^{m\times m}$ 和 $\operatorname{rank}(AB) \leqslant n < m$ 立即导出 $\det AB = 0$。

以下假设 $m \leqslant n$。显然

$$AB = \begin{pmatrix} \sum_{i_1=1}^{n} \alpha_{1i_1}\beta_{i_1 1} \cdots \sum_{i_m=1}^{n} \alpha_{1i_m}\beta_{i_m m} \\ \vdots \\ \sum_{i_1=1}^{n} \alpha_{mi_1}\beta_{i_1 1} \cdots \sum_{i_m=1}^{n} \alpha_{mi_m}\beta_{i_m m} \end{pmatrix}.$$

利用行列式的运算法则，

$$\det AB = \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} \begin{vmatrix} \alpha_{1i_1}\beta_{i_1 1} \cdots \alpha_{1i_m}\beta_{i_m m} \\ \vdots \\ \alpha_{mi_1}\beta_{i_1 1} \cdots \alpha_{mi_m}\beta_{i_m m} \end{vmatrix}$$

$$= \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} A\begin{pmatrix} 1 \cdots m \\ i_1 \cdots i_m \end{pmatrix} \beta_{i_1 1} \cdots \beta_{i_m m}. \tag{8.2}$$

在(8.2)式右端共有 n^m 项. 但注意到，当 $i_1, \cdots, i_m$ 中至少有两个指标相等时，必有 $A\begin{pmatrix} 1 \cdots m \\ i_1 \cdots i_m \end{pmatrix} = 0$，所以只需考虑

$$C_n^m \cdot m! = \frac{n!}{(n-m)!}$$

项，其中每一项所含的指标 $i_1, \cdots, i_m$ 互不相等. 因此

$$\det AB = \sum_{\substack{i_1, \cdots, i_m = 1 \\ i_s \neq i_t (s \neq t), s, t = 1, \cdots, m}}^{n} A\begin{pmatrix} 1 \cdots m \\ i_1 \cdots i_m \end{pmatrix} \beta_{i_1 1} \cdots \beta_{i_m m}. \tag{8.3}$$

现在把 (8.3) 式右端的 $\frac{n!}{(n-m)!}$ 项分成 C_m^n 组，每一组内的各项仅仅是指标 $i_1, \cdots, i_m$ 的排列次序不同. 于是(8.3)式右端对应于 $1, \cdots, n$ 中任意固定的 m 个数 $k_1 < \cdots < k_m$ 者，共有 $m!$ 项，它们构成一组，其代数和为

$$g_{k_1, \cdots, k_m} = \sum_{i_1, \cdots, i_m}^{(k)} A\begin{pmatrix} 1 \cdots m \\ i_1 \cdots i_m \end{pmatrix} \beta_{i_1 1} \cdots \beta_{i_m m}.$$

这里的和号 $\sum\limits_{i_1,\cdots,i_m}^{(k)}$ 表示 $i_1,\cdots,i_m$ 取遍 $k_1,\cdots,k_m$ 的所有可能的排列。

用 $[i_1\cdots i_m]$ 表示 $i_1,\cdots,i_m$ 按所述顺序的一个排列，$[i_1\cdots i_m]$中一对指标的互换叫做一次对换，$t\begin{bmatrix} i_1 & \cdots i_m \\ k_1 & \cdots k_m \end{bmatrix}$表示由$[i_1\cdots i_m]$变为 $[k_1\cdots k_m]$ 所需对换的次数，从行列式的定义得到

$$
\begin{aligned}
g_{k_1,\cdots,k_m} &= \sum_{i_1,\cdots,i_m}^{(k)} (-1)^{t\left[\begin{smallmatrix} i_1\cdots i_m \\ k_1\cdots k_m \end{smallmatrix}\right]} A\begin{pmatrix} 1\cdots m \\ k_1\cdots k_m \end{pmatrix} \beta_{i_1 1}\cdots\beta_{i_m m} \\
&= A\begin{pmatrix} 1\cdots m \\ k_1\cdots k_m \end{pmatrix} \sum_{i_1,\cdots,i_m}^{(k)} (-1)^{t\left[\begin{smallmatrix} i_1\cdots i_m \\ k_1\cdots k_m \end{smallmatrix}\right]} \beta_{i_1 1}\cdots\beta_{i_m m} \\
&= A\begin{pmatrix} 1\cdots m \\ k_1\cdots k_m \end{pmatrix} B\begin{pmatrix} k_1\cdots k_m \\ 1\cdots m \end{pmatrix}.
\end{aligned}
$$

再结合(8.3)，立即导出(8.1)中当 $m \leqslant n$ 时的等式。□

8.2. Hadamard 不等式

定义 8.1. 设 $A = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m)$, $\boldsymbol{a}_i \in \mathbf{C}^n$, $i = 1, \cdots, m$. 则

$$G(A) = \det A^H A \tag{8.4}$$

叫做向量 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m$ 的 Gram 行列式（或 A 的 Gram 行列式）。

容易看出，对于 $A \in \mathbf{C}^{n\times m}$，当 $m > n$ 时，必有 $G(A) = 0$.

设 $A_l \in \mathbf{C}^{n\times l} (1 \leqslant l < n)$. 又设 $\mathbf{C}^n$ 中的向量 $\boldsymbol{a}$ 关于 $R(A_l)$ 和 $R(A_l)^\perp$ 的唯一分解（见(7.5)式）为

$$
\begin{gathered}
\boldsymbol{a} = \boldsymbol{r} + \hat{\boldsymbol{n}}, \boldsymbol{r} = P_{R(A_l)}\boldsymbol{a} \in R(A_l), \\
\hat{\boldsymbol{n}} = P_{R(A_l)^\perp}\boldsymbol{a} \in R(A_l)^\perp.
\end{gathered}
$$

由

$$\|\boldsymbol{a}\|_2^2 = \|\boldsymbol{r}\|_2^2 + \|\hat{\boldsymbol{n}}\|_2^2$$

知

$$0 \leqslant \frac{\|\hat{\boldsymbol{n}}\|_2}{\|\boldsymbol{a}\|_2} \leqslant 1.$$

因此可记

$$\sin\theta(\boldsymbol{a},R(A_l))=\frac{\|\hat{\boldsymbol{n}}\|_2}{\|\boldsymbol{a}\|_2},$$

$$0\leqslant\theta(\boldsymbol{a},R(A_l))\leqslant\frac{\pi}{2}. \tag{8.5}$$

我们从平面几何学知道，如果 $\boldsymbol{a}_1$ 与 $\boldsymbol{a}$ 是 $\mathbf{R}^2$ 中的任二向量(见图 1-4)，则

$$G((\boldsymbol{a}_1,\boldsymbol{a}))=\begin{vmatrix}\boldsymbol{a}_1^T\boldsymbol{a}_1 & \boldsymbol{a}_1^T\boldsymbol{a}\\ \boldsymbol{a}^T\boldsymbol{a}_1 & \boldsymbol{a}^T\boldsymbol{a}\end{vmatrix}=\boldsymbol{a}_1^T\boldsymbol{a}_1\boldsymbol{a}^T\boldsymbol{a}-(\boldsymbol{a}_1^T\boldsymbol{a})^2$$

$$=\|\boldsymbol{a}_1\|_2^2\|\boldsymbol{a}\|_2^2(1-\cos^2\theta(\boldsymbol{a},R(\boldsymbol{a}_1)))$$

$$=\|\boldsymbol{a}\|_2^2\|\boldsymbol{a}_1\|_2^2\sin^2\theta(\boldsymbol{a},R(\boldsymbol{a}_1)).$$

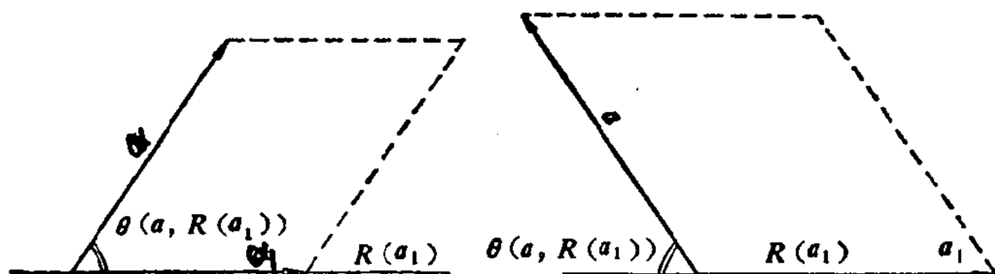


图 1-4

下面的引理 8.1 推广了这一重要结论.

引理 8.1. 设 $A_l=(\boldsymbol{a}_1,\cdots,\boldsymbol{a}_l)$，$\boldsymbol{a}_i\in\mathbf{C}^n$，$i=1,\cdots,l$. 对于任一向量 $\boldsymbol{a}\in\mathbf{C}^n$，如果记 $A=(A_l,\boldsymbol{a})$，则

$$G(A)=\|\boldsymbol{a}\|_2^2G(A_l)\sin^2\theta(\boldsymbol{a},R(A_l)), \tag{8.6}$$

其中 $G(A)$ 与 $G(A_l)$ 分别表示 A 与 A_l 的 Gram 行列式，$\sin\theta(\boldsymbol{a},R(A_l))$ 的定义见(8.5).

证明:

首先注意到，当 $\operatorname{rank}A_l<l$ 时，(8.6)式两端均取零值，这时(8.6)式显然成立；当 $\operatorname{rank}A_l=l=n$ 时，

$$G(A)=\sin\theta(\boldsymbol{a},R(A_l))=0,$$

(8.6)式亦显然成立. 所以，只需讨论 $l<n$ 并且 $A_l\in\mathbf{C}_l^{n\times l}$ 的情形.

将 $\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_n)^T$ 分解为

$$\left.\begin{aligned}\boldsymbol{\alpha}=\boldsymbol{r}+\hat{\boldsymbol{n}},\quad \boldsymbol{r}=(\rho_1,\cdots,\rho_n)^T\in R(A_l),\\ \hat{\boldsymbol{n}}=(\nu_1,\cdots,\nu_n)^T\in R(A_l)^{\perp}.\end{aligned}\right\}\tag{8.7}$$

由 $\boldsymbol{r}=A_l\boldsymbol{c}$，$\boldsymbol{c}\in\mathbf{C}^l$ 以及 $A_l^H\hat{\boldsymbol{n}}=0$ 得出

$$A_l^H(\boldsymbol{\alpha}-A_l\boldsymbol{c})=0,$$

因而有

$$\begin{cases}A_l^HA_l\boldsymbol{c}+A_l^H\boldsymbol{\alpha}(-1)=0\\ A_l\boldsymbol{c}+\boldsymbol{r}(-1)=0.\end{cases}\tag{8.8}$$

记 $A_l=\begin{pmatrix}\boldsymbol{\alpha}^{(1)}\\ \vdots\\ \boldsymbol{\alpha}^{(n)}\end{pmatrix}$，$\boldsymbol{\alpha}^{(i)T}\in\mathbf{C}^l$，$i=1,\cdots,n$. 则(8.8)式可写成

$$\begin{pmatrix}A_l^HA_l & A_l^H\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{(i)} & \rho_i\end{pmatrix}\begin{pmatrix}\boldsymbol{c}\\ -1\end{pmatrix}=0,\ i=1,\cdots,n,$$

由此立即得到

$$\det\begin{pmatrix}A_l^HA_l & A_l^H\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{(i)} & \rho_i\end{pmatrix}=0,\ i=1,\cdots,n.\tag{8.9}$$

考虑到

$$\begin{aligned}\det\begin{pmatrix}A_l^HA_l & A_l^H\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{(i)} & \rho_i\end{pmatrix}&=\det\begin{pmatrix}A_l^HA_l & A_l^H\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{(i)} & 0\end{pmatrix}+\det\begin{pmatrix}A_l^HA_l & 0\\ \boldsymbol{\alpha}^{(i)} & \rho_i\end{pmatrix}\\ &=\det\begin{pmatrix}A_l^HA_l & A_l^H\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{(i)} & 0\end{pmatrix}+\rho_iG(A_l),\quad i=1,\cdots,n.\end{aligned}$$

所以从(8.9)可解出

$$\rho_i=-\frac{\det\begin{pmatrix}A_l^HA_l & A_l^H\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{(i)} & 0\end{pmatrix}}{G(A_l)},\ i=1,\cdots,n.$$

代入(8.7)，得到

$$\nu_i=\alpha_i-\rho_i=\frac{\det\begin{pmatrix}A_l^HA_l & A_l^H\boldsymbol{\alpha}\\ \boldsymbol{\alpha}^{(i)} & \alpha_i\end{pmatrix}}{G(A_l)},\ i=1,\cdots,n.$$

因此

$$\|\hat{\boldsymbol{n}}\|_2^2 = \hat{\boldsymbol{n}}^H\hat{\boldsymbol{n}} = \boldsymbol{\alpha}^H\hat{\boldsymbol{n}} = \sum_{i=1}^{n} \bar{\alpha}_i\nu_i$$

$$= \sum_{i=1}^{n} \det\begin{pmatrix} A_l^H A_l & A_l^H\boldsymbol{a} \\ \bar{\alpha}_i\boldsymbol{a}^{(i)} & \bar{\alpha}_i\alpha_i \end{pmatrix} \Big/ G(A_l)$$

$$= \frac{\det\begin{pmatrix} A_l^H A_l & A_l^H\boldsymbol{a} \\ \boldsymbol{a}^H A_l & \boldsymbol{a}^H\boldsymbol{a} \end{pmatrix}}{G(A_l)} = \frac{G(A)}{G(A_l)}.$$

再联系到(8.5),便导出

$$\|\boldsymbol{a}\|_2^2\sin^2\theta(\boldsymbol{a}, R(A_l)) = \frac{G(A)}{G(A_l)},$$

即(8.6)式成立. □

利用引理 8.1,通过递推可得

定理 8.2. 设 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_l \in \mathbb{C}^n$, $A_k = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k)$, $k = 1, \cdots, l$. 则

$$G(A_l) = \prod_{i=1}^{l} \|\boldsymbol{a}_i\|_2^2 \cdot \prod_{i=2}^{l} \sin^2\theta(\boldsymbol{a}_i, R(A_{i-1})). \tag{8.10}$$

作为定理 8.2 的推论,我们得到

定理 8.3 (Hadamard 不等式). 设 $A = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_l)$, $\boldsymbol{a}_i \in \mathbb{C}^n, i = 1, \cdots, l$. 则

$$G(A) \leqslant \prod_{i=1}^{l} \|\boldsymbol{a}_i\|_2^2. \tag{8.11}$$

等式成立的充要条件是 $a_i^T a_j = 0$, $i \neq j$, $i, j = 1, \cdots, n$.

下面的定理 8.4 和定理 8.5,与 Hadamard 不等式等价.

定理 8.4. 设 $A = (\alpha_{ij}) \in \mathbb{C}^{n\times n}$. 则有

$$|\det A|^2 \leqslant \prod_{j=1}^{n}\left(\sum_{i=1}^{n} |\alpha_{ij}|^2\right) \tag{8.12}$$

和

$$|\det A|^2 \leqslant \prod_{i=1}^{n}\left(\sum_{j=1}^{n} |\alpha_{ij}|^2\right). \tag{8.13}$$

定理 8.5. 设 $A = (\alpha_{ij}) \in \mathbb{C}^{n\times n}$ 为半正定阵. 则

$$\det A \leqslant \prod_{i=1}^{n} \alpha_{ii}. \tag{8.14}$$

注 8.1. Hadamard 不等式已经有 90 年的历史了. 90 年来,积累了许多不同的证法. 本节的证明方法,是作者参考 Гантмахер [197] 给出的. 对 Hadamard 不等式有兴趣的读者,可参阅 E. F. Beckenbach 与 R. Bellman 的专著[52]. 对于定理 8.5,下面再给出一个证明:

无妨假设 A 为正定阵(因为当 A 为半正定阵时,可首先考虑正定阵 $A+\varepsilon I$, 其中 $\varepsilon>0$; 然后取极限 $\varepsilon\to 0$ 即可). 由定理 3.2 知, A 有三角分解

$$A=\begin{pmatrix} l_{11} & 0 \\ \boldsymbol{l} & L \end{pmatrix}\begin{pmatrix} \bar{l}_{11} & \boldsymbol{l}^H \\ 0 & L^H \end{pmatrix},\ L\in \mathbf{C}^{(n-1)\times(n-1)}. \tag{8.15}$$

上式亦可改写成

$$\begin{pmatrix} \alpha_{11} & \boldsymbol{a}^H \\ \boldsymbol{a} & A_1 \end{pmatrix}=\begin{pmatrix} |l_{11}|^2 & * \\ * & \boldsymbol{l}\boldsymbol{l}^H+LL^H \end{pmatrix}.$$

在(8.15)式两端同时取行列式,得到

$$\begin{aligned}\det A &= |l_{11}|^2|\det L|^2 \\ &\leqslant |l_{11}|^2\det(\boldsymbol{l}\boldsymbol{l}^H+LL^H)=\alpha_{11}\det A_1.\end{aligned}$$

再对正定阵 A_1 进行同样的讨论,依次类推即可得到不等式(8.14). □

习题

1. 设 $\boldsymbol{a}=(\alpha_1,\cdots,\alpha_n)^T$ 与 $\boldsymbol{b}=(\beta_1,\cdots,\beta_n)^T\in\mathbf{C}^n$. 试利用 Binet-Cauchy 公式证明 Cauchy 不等式

$$|\boldsymbol{a}^H\boldsymbol{b}|\leqslant\|\boldsymbol{a}\|_2\|\boldsymbol{b}\|_2.$$

2. 设 A 与 B 为半正定阵. 则

$$\det(A+B)\geqslant\det A+\det B.$$

3. 设 $A=(\alpha_{ij})\in\mathbf{C}^{n\times n}$, $\alpha=\max\limits_{i,j}|\alpha_{ij}|$, 试证

$$|\det A|\leqslant n^{\frac{n}{2}}\alpha^n.$$

4. 设 $A=\begin{pmatrix} A_{11}\cdots A_{1k} \\ \vdots \qquad \vdots \\ A_{k1}\cdots A_{kk}\end{pmatrix}$ 为半正定阵,其中 $A_{ii}(1\leqslant i\leqslant k)$均

为方阵．试证

$$\det A \leqslant \prod_{i=1}^{k} \det A_{ii}$$

（推广的 Hadamard 不等式）．

5．设 $H = \begin{pmatrix} A & B \\ B^H & C \end{pmatrix}$ 为正定阵，其中 $B \in \mathbb{C}^{m\times n}$．试证：对于任何 $R \in \mathbb{C}^{m\times n}$，恒有

$$\det H \leqslant \det A \cdot \det(C + R^H B + B^H R + R^H A R),$$

并且等号成立的必要与充分条件是 $AR + B = 0$．

第一章说明

这一章的几乎每一节都是独立的课题，并且已有相当丰富的积累．本章所论述的，仅仅是与矩阵扰动分析有关的一些基本概念和重要结果，主要参考 Householder [106]、Ben-Israel 与 Greville [55]、Гантмахер [197] 和 Stewart [157]．

第二章 范数与度量

§1 $\mathbb{C}^n$ 上的范数

$\mathbf{C}^n$ 上的范数的概念是复数的模的概念的推广. 复数 ξ 的模 $|\xi|$ 是 $\mathbf{C}$ 到 $\mathbb{R}_+$ (非负实数集合)的一个函数,它有三条性质:

1) $\xi \neq 0 \Longrightarrow |\xi| > 0$ (正定性),

2) $|\alpha\xi| = |\alpha||\xi|$ (齐次性),

3) $|\xi + \eta| \leqslant |\xi| + |\eta|$ (三角不等式).

定义 1.1. 一个从 $\mathbf{C}^n$ 到 $\mathbf{R}_+$ 的函数 ν 叫做 $\mathbf{C}^n$ 上的范数,如果它满足

1) $\boldsymbol{x} \neq 0 \Longrightarrow \nu(\boldsymbol{x}) > 0$ (正定性),

2) $\nu(\alpha\boldsymbol{x}) = |\alpha|\nu(\boldsymbol{x})$ (齐次性),

3) $\nu(\boldsymbol{x} + \boldsymbol{y}) \leqslant \nu(\boldsymbol{x}) + \nu(\boldsymbol{y})$ (三角不等式).

其中 $\boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T$ 与 $\boldsymbol{y} = (\eta_1, \cdots, \eta_n)^T$ 是 $\mathbf{C}^n$ 中任意的向量,α 是任意的复数.

从定义 1.1 容易导出

$$\nu(0) = 0, \nu(-\boldsymbol{x}) = \nu(\boldsymbol{x})$$

和

$$|\nu(\boldsymbol{x}) - \nu(\boldsymbol{y})| \leqslant \nu(\boldsymbol{x} - \boldsymbol{y}). \tag{1.1}$$

$\mathbf{C}^n$ 上的下列几种范数是常用的:

例 1.1. $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |\xi_i|, \forall \boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T \in \mathbf{C}^n$.

例 1. 2. $\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{n} |\xi_i|^2}, \forall \boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T \in \mathbf{C}^n$.

例 1.3. $\|\boldsymbol{x}\|_\infty = \max_{1 \leqslant i \leqslant n} |\xi_i|, \forall \boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T \in \mathbf{C}^n$.

上述三种范数,可以看作下述 Hölder 范数(或 p 范数)的特

殊情形:

例 1.4. $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |\xi_i|^p\right)^{\frac{1}{p}}, 1 \leqslant p < \infty, \forall \boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T \in \mathbf{C}^n$.

本节习题 2 的提示,将指出证明 $\|\boldsymbol{x}\|_p$ 是范数的一种方法.

当 $p = 1,2$ 时, $\|\boldsymbol{x}\|_p$ 分别是 $\|\boldsymbol{x}\|_1$ 和 $\|\boldsymbol{x}\|_2$. 此外,有

$$\lim_{p\to\infty} \|\boldsymbol{x}\|_p = \|\boldsymbol{x}\|_\infty.$$

下述定理指出,可以利用已知的范数去构造新范数.

定理 1.1. 设 μ 是 $\mathbf{C}^m$ 上的范数, $A \in \mathbf{C}_n^{m\times n}$. 则由

$$\nu(\boldsymbol{x}) = \mu(A\boldsymbol{x}), \ \boldsymbol{x} \in \mathbf{C}^n$$

所定义的 ν 是 $\mathbf{C}^n$ 上的范数.

证明:

只需验证 ν 满足定义 1.1 中的 1), 2)与 3).

1) $\boldsymbol{x} \neq 0 \Longrightarrow A\boldsymbol{x} \neq 0 \Longrightarrow \nu(\boldsymbol{x}) = \mu(A\boldsymbol{x}) > 0$,

2) $\nu(\alpha\boldsymbol{x}) = \mu(\alpha A\boldsymbol{x}) = |\alpha|\mu(A\boldsymbol{x}) = |\alpha|\nu(\boldsymbol{x})$,

3) $\nu(\boldsymbol{x}+\boldsymbol{y}) = \mu(A(\boldsymbol{x}+\boldsymbol{y})) = \mu(A\boldsymbol{x}+A\boldsymbol{y})$

$\leqslant \mu(A\boldsymbol{x}) + \mu(A\boldsymbol{y}) = \nu(\boldsymbol{x}) + \nu(\boldsymbol{y})$. □

推论 1.1. 设 $A \in \mathbf{C}^{n\times n}$ 为正定阵. 则由

$$\nu(\boldsymbol{x}) = \sqrt{\boldsymbol{x}^H A\boldsymbol{x}}, \ \boldsymbol{x} \in \mathbf{C}^n.$$

定义的 ν 是 $\mathbf{C}^n$ 上的范数.

证明:

根据正定阵的三角分解(第一章定理 3.2), $A = LL^H$, L 是具有正对角线元素的下三角阵. 于是

$$\nu(\boldsymbol{x}) = \|L^H\boldsymbol{x}\|_2.$$

再由定理 1.1 可知 ν 是 $\mathbf{C}^n$ 上的范数. □

通常在 $\mathbf{C}^n$ 上以$\|\ \|_2$为范数. 由$\|\ \|_2$在 $\mathbf{C}^n$ 诱导出的度量,叫做 Euclid 度量, 即对于 $\mathbf{C}^n$ 内的任意二点 $\boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T$ 与 $\boldsymbol{y} = (\eta_1, \cdots, \eta_n)^T$,用

$$\|\boldsymbol{x}-\boldsymbol{y}\|_2 = \sqrt{\sum_{i=1}^{n} |\xi_i - \eta_i|^2}$$

作为它们之间的距离，从而 $\mathbf{C}^n$ 成了一个度量空间．下面就是在这种度量的意义下讨论问题(比如开集，闭集，极限，连续性，等等)．

引理 1.1. 设 ν 是 $\mathbf{C}^n$ 上的范数，则 $\nu(\boldsymbol{x})$ 是 $\boldsymbol{x}\in\mathbf{C}^n$ 的连续函数．

证明:

按照连续函数的定义，只需证明：任给 $\varepsilon>0$，存在 $\delta>0$，使当 $\|\boldsymbol{y}-\boldsymbol{x}\|_2<\delta$ 时，有 $|\nu(\boldsymbol{y})-\nu(\boldsymbol{x})|<\varepsilon$，此处 $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T$ 与 $\boldsymbol{y}=(\eta_1,\cdots,\eta_n)^T\in\mathbf{C}^n$.

今以 $\boldsymbol{e}_i$ 表示 $I^{(n)}$ 的第 i 列．利用 (1.1) 和范数的性质 2) 与 3)，可得

$$
\begin{aligned}
|\nu(\boldsymbol{y})-\nu(\boldsymbol{x})| &\leqslant \nu(\boldsymbol{y}-\boldsymbol{x})=\nu\left(\sum_{i=1}^{n}(\eta_i-\xi_i)\boldsymbol{e}_i\right)\\
&\leqslant \sum_{i=1}^{n}|\eta_i-\xi_i|\,\nu(\boldsymbol{e}_i)\leqslant\sqrt{\sum_{i=1}^{n}|\eta_i-\xi_i|^2}\\
&\quad\times\sqrt{\sum_{i=1}^{n}\nu^2(\boldsymbol{e}_i)}=\gamma\|\boldsymbol{y}-\boldsymbol{x}\|_2,
\end{aligned}
$$

其中 $\gamma=\sqrt{\sum_{i=1}^{n}\nu^2(\boldsymbol{e}_i)}>0$，且与 x，y 无关．显然只需取 $\delta=\dfrac{\varepsilon}{\gamma}$，则当 $\|\boldsymbol{y}-\boldsymbol{x}\|_2<\delta$ 时，必有 $|\nu(\boldsymbol{y})-\nu(\boldsymbol{x})|<\varepsilon$. □

引理 1.2. 点集

$$
\mathscr{S}_\infty=\{\boldsymbol{x}\in\mathbf{C}^n:\|\boldsymbol{x}\|_\infty=1\} \tag{1.2}
$$

是 $\mathbf{C}^n$ 内一有界闭集．

证明:

对于任一点 $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T\in\mathscr{S}_\infty$，由

$$
\|\boldsymbol{x}\|_2=\sqrt{\sum_{i=1}^{n}|\xi_i|^2}\leqslant\sqrt{n}\;\|\boldsymbol{x}\|_\infty=\sqrt{n}
$$

可知，$\mathscr{S}_\infty$ 是 $\mathbf{C}^n$ 内一有界集．

为了证明 $\mathscr{S}_\infty$ 是 $\mathbf{C}^n$ 内一闭集，只需证明：如果 $\{\boldsymbol{x}_k\}$ 是 $\mathscr{S}_\infty$

内一无穷点列，并且 $\lim\limits_{k\to\infty}\|\boldsymbol{x}_k-\boldsymbol{x}\|_2=0$，则 $\boldsymbol{x}\in\mathscr{S}_\infty$.

显然存在 $\{\boldsymbol{x}_k\}$ 的一个无穷子列 $\{\boldsymbol{x}_{k_i}\}$，使得对于某一个确定的 $j(1\leqslant j\leqslant n)$，$\boldsymbol{x}_{k_i}$ 的第 j 个分量 $\xi_j^{(k_i)}$ 的模为 1，$i=1,2,\cdots$. 设 $\boldsymbol{x}$ 的第 j 个分量为 ξ_j，则由

$$|1-|\xi_j||=||\xi_j^{(k_i)}|-|\xi_j||\leqslant|\xi_j^{(k_i)}-\xi_j|\leqslant\|\boldsymbol{x}_{k_i}-\boldsymbol{x}\|_2\to 0 \quad (i\to\infty)$$

立即得到 $|\xi_j|=1$，即 $\boldsymbol{x}\in\mathscr{S}_\infty$. □

利用引理 1.1 和引理 1.2，可证下述关于 $\mathbb{C}^n$ 上的范数的等价性定理.

定理 1.2. 设 ν 与 μ 是 $\mathbb{C}^n$ 上的范数. 则存在仅与 ν,μ 有关的正数 r_1 与 r_2，使得

$$r_1\nu(\boldsymbol{x})\leqslant\mu(\boldsymbol{x})\leqslant r_2\nu(\boldsymbol{x}),\ \forall\boldsymbol{x}\in\mathbb{C}^n. \tag{1.3}$$

证明:

当 $\boldsymbol{x}=0$ 时，(1.3)显然成立. 所以以下考虑任一 $\boldsymbol{x}\in\mathbb{C}^n$，$\boldsymbol{x}\neq 0$.

一方面，有

$$\mu(\boldsymbol{x})=\mu\left(\sum_{i=1}^n\xi_i\boldsymbol{e}_i\right)\leqslant\sum_{i=1}^n|\xi_i|\mu(\boldsymbol{e}_i)\leqslant\|\boldsymbol{x}\|_\infty p_2, \tag{1.4}$$

其中 $p_2=\sum\limits_{i=1}^n\mu(\boldsymbol{e}_i)>0$ 与 $\boldsymbol{x}$ 无关.

另一方面，据引理 1.1，μ 是 $\mathbb{C}^n$ 上的连续函数；据引理 1.2，(1.2)所示的 $\mathscr{S}_\infty$ 是 $\mathbb{C}^n$ 内一有界闭集. 因此，μ 必在 $\mathscr{S}_\infty$ 上达到极小值 $p_1=\min\limits_{\boldsymbol{y}\in\mathscr{S}_\infty}\mu(\boldsymbol{y})>0$. 现令 $\boldsymbol{x}=\|\boldsymbol{x}\|_\infty\boldsymbol{y}$，显然 $\boldsymbol{y}\in\mathscr{S}_\infty$. 因此

$$\mu(\boldsymbol{x})=\mu(\|\boldsymbol{x}\|_\infty\boldsymbol{y})=\|\boldsymbol{x}\|_\infty\mu(\boldsymbol{y})\geqslant\|\boldsymbol{x}\|_\infty p_1. \tag{1.5}$$

(1.4)与(1.5)给出

$$\|\boldsymbol{x}\|_\infty p_1\leqslant\mu(\boldsymbol{x})\leqslant\|\boldsymbol{x}\|_\infty p_2, \tag{1.6}$$

其中正数 p_1，p_2 与 $\boldsymbol{x}$ 无关.

同理可得

$$\|\boldsymbol{x}\|_\infty q_1 \leqslant \nu(\boldsymbol{x}) \leqslant \|\boldsymbol{x}\|_\infty q_2, \tag{1.7}$$

其中正数 q_1, q_2 亦与 $\boldsymbol{x}$ 无关.

由(1.6)和(1.7)立即得出

$$0 < r_1 = \frac{p_1}{q_2} \leqslant \frac{\mu(\boldsymbol{x})}{\nu(\boldsymbol{x})} \leqslant \frac{p_2}{q_1} = r_2. \quad \square$$

例 1.5.

$$\|\boldsymbol{x}\|_\infty \leqslant \|\boldsymbol{x}\|_1 \leqslant n\|\boldsymbol{x}\|_\infty,$$

$$\frac{1}{\sqrt{n}} \|\boldsymbol{x}\|_1 \leqslant \|\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{x}\|_1,$$

$$\frac{1}{\sqrt{n}} \|\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{x}\|_\infty \leqslant \|\boldsymbol{x}\|_2.$$

由 $\mathbf{C}^n$ 上的不同范数，可以诱导出 $\mathbf{C}^n$ 上的不同度量，从而在 $\mathbf{C}^n$ 上建立不同的拓扑. 定理 1.2 表明，这些不同的拓扑是等价的. 比如，下面的定理 1.3 指出了极限定义的等价性.

设 $\boldsymbol{x}_k = (\xi_1^{(k)}, \cdots, \xi_n^{(k)})^T \in \mathbf{C}^n$, $k = 1, 2, \cdots$. 如果

$$\lim_{k\to\infty} \xi_i^{(k)} = \xi_i \quad (i = 1, \cdots, n),$$

则称向量序列 $\{\boldsymbol{x}_k\}$ 有极限 $\boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T \in \mathbf{C}^n$，或称 $\boldsymbol{x}_k$ 收敛到 $\boldsymbol{x}$，记作

$$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}.$$

定理 1.3. 设 ν 是 $\mathbf{C}^n$ 上的任一范数， $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \in \mathbf{C}^n$. 则

$$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x} \Longleftrightarrow \lim_{k\to\infty} \nu(\boldsymbol{x}_k - \boldsymbol{x}) = 0. \tag{1.8}$$

证明:

记 $\boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T$, $\boldsymbol{x}_k = (\xi_1^{(k)}, \cdots, \xi_n^{(k)})^T$, $k = 1, 2, \cdots$. 首先容易得出

$$\lim_{k\to\infty} \xi_i^{(k)} = \xi_i (i = 1, \cdots, n) \Longleftrightarrow \lim_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}\|_\infty = 0; \tag{1.9}$$

再利用定理 1.2，存在与 $\boldsymbol{x}, \boldsymbol{x}_k (k = 1, 2, \cdots)$ 无关的正数 s_1 与 s_2，使得

$$s_1\nu(\boldsymbol{x}_k - \boldsymbol{x}) \leqslant \|\boldsymbol{x}_k - \boldsymbol{x}\|_\infty \leqslant s_2\nu(\boldsymbol{x}_k - \boldsymbol{x}),$$

因而

$$\lim_{k\to\infty}\|\boldsymbol{x}_k-\boldsymbol{x}\|_\infty=0\Longleftrightarrow\lim_{k\to\infty}\nu(\boldsymbol{x}_k-\boldsymbol{x})=0. \qquad (1.10)$$

把(1.9)与(1.10)联系起来,便得到(1.8). □

习题

1. 证明$\|\ \|_1$, $\|\ \|_2$和$\|\ \|_\infty$满足三角不等式.

2. 证明$\|\ \|_p$满足三角不等式(提示如下:

① Young (1921) 曾经证明: 若当 $\xi\geqslant 0$ 时, $\eta=\varphi(\xi)$是严格单调递增函数,且满足 $\varphi(0)=0$, $\lim\limits_{\xi\to+\infty}\varphi(\xi)=+\infty$. 则对于任意二正数$\alpha$与$\beta$,有

$$\alpha\beta\leqslant\int_0^\alpha\varphi(\xi)d\xi+\int_0^\beta\varphi^{-1}(\eta)d\eta. \qquad (*)$$

利用 Young 不等式(*)证明: 对于$\alpha,\beta>0$与$p>1$,有

$$\alpha\beta\leqslant\frac{\alpha^p}{p}+\frac{\beta^q}{q},\ \frac{1}{p}+\frac{1}{q}=1. \qquad (**)$$

② 利用不等式(**)证明: 若$\boldsymbol{x},\boldsymbol{y}\in\mathbf{C}^n$,则

$$|\boldsymbol{x}^H\boldsymbol{y}|\leqslant\|\boldsymbol{x}\|_p\|\boldsymbol{y}\|_q,\ \frac{1}{p}+\frac{1}{q}=1. \qquad (***)$$

③ 利用不等式(***)证明

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p\leqslant\|\boldsymbol{x}\|_p+\|\boldsymbol{y}\|_p.\ )$$

3. 设ν是$\mathbf{C}^n$上的任一范数. 试证

$$\mathscr{I}_\nu=\{\boldsymbol{x}\in\mathbf{C}^n:\nu(\boldsymbol{x})=1\}$$

是$\mathbf{C}^n$内一有界闭集.

4. 设ν是$\mathbf{C}^n$上的任一范数. 令

$$\mathscr{B}_\nu=\{\boldsymbol{x}\in\mathbf{C}^n:\nu(\boldsymbol{x})\leqslant 1\}.$$

试证$\mathscr{B}_\nu$是一个均衡凸体(包含内点的、均衡的有界闭凸集),即证明:

(i) 存在 $\varepsilon>0$, 使得 $\{\boldsymbol{x}\in\mathbf{C}^n:\|\boldsymbol{x}\|_2\leqslant\varepsilon\}\subset\mathscr{B}_\nu$,

(ii) 对一切满足$|\alpha|\leqslant 1$的$\alpha\in\mathbf{C}$,必有$\alpha\boldsymbol{x}\in\mathscr{B}_\nu$, $\forall\boldsymbol{x}\in\mathscr{B}_\nu$,

(iii) $\mathscr{B}_\nu$在$\mathbf{C}^n$内有界,

(iv) $\mathscr{B}_\nu$ 是 $\mathbf{C}^n$ 内的闭集,

(v) $\mathscr{B}_\nu$ 是 $\mathbf{C}^n$ 内的凸集.

§2 $\mathbf{C}^{m\times n}$ 上的范数

2.1 基本概念

把 $\mathbf{C}^n$ 上的范数概念推广到 $\mathbf{C}^{m\times n}$ 上,便是 $\mathbf{C}^{m\times n}$ 上的范数.

定义 2.1. 一个从 $\mathbf{C}^{m\times n}$ 到 $\mathbf{R}_+$ 上的函数 ν 叫做 $\mathbf{C}^{m\times n}$ 上的范数,如果它满足

I) $A \neq 0 \Longrightarrow \nu(A) > 0$ (正定性),

II) $\nu(\alpha A) = |\alpha|\nu(A)$ (齐次性),

III) $\nu(A + B) \leqslant \nu(A) + \nu(B)$ (三角不等式).

其中 A, B 是 $\mathbf{C}^{m\times n}$ 中任意的矩阵, α 是任意的复数.

从定义 2.1 容易导出

$$\nu(0) = 0, \quad \nu(-A) = \nu(A)$$

和

$$|\nu(A) - \nu(B)| \leqslant \nu(A - B).$$

例 2.1. 对于 $A = (\alpha_{ij}) \in \mathbf{C}^{m\times n}$, 令

$$\|A\|_F = \sqrt{\sum_{i,j} |\alpha_{ij}|^2}.$$

$\|A\|_F$叫做 A 的 Frobenius 范数(或者叫做 A 的 Euclid 范数),它是把 A 视为 $\mathbf{C}^{mn}$ 中的一个向量,取其 Euclid 长度所得到的范数. 因此,矩阵的 Frobenius 范数是向量的 Euclid 范数的自然推广.

例 2.2. 对于 $A = (\alpha_{ij}) \in \mathbf{C}^{m\times n}$, 令

$$\|A\|'_\infty = \max_{i,j} |\alpha_{ij}|.$$

范数 $\|\ \ \|_\infty'$ 是向量范数 $\|\ \ \|_\infty$ 的自然推广.

例 2.3. 对于 $A = (\alpha_{ij}) \in \mathbf{C}^{n\times n}$, 令

$$\|A\|_\alpha = \frac{1}{n} \sum_{i,j=1}^{n} |\alpha_{ij}|.$$

范数$\| \quad \|_a$是向量范数$\| \quad \|_1$的一个推广.

注意到，如果把 $A \in \mathbf{C}^{m\times n}$ 看作 $\mathbf{C}^{mn}$ 中的一个向量，则 $\mathbf{C}^{m\times n}$ 上的任一矩阵范数，可以看作 $\mathbf{C}^{mn}$ 上的一个向量范数. 根据定理 1.2，立即得出

定理 2.1. 设 ν 与 μ 是 $\mathbf{C}^{m\times n}$ 上的范数. 则存在仅与 ν, μ 有关的正数 s_1 与 s_2，使得

$$s_1\nu(A) \leqslant \mu(A) \leqslant s_2\nu(A), \quad \forall A \in \mathbf{C}^{m\times n}.$$

设 $A_k = (\alpha_{ij}^{(k)}) \in \mathbf{C}^{m\times n}$, $k = 1, 2, \cdots$. 如果

$$\lim_{k\to\infty} \alpha_{ij}^{(k)} = \alpha_{ij} (i = 1, \cdots, m; j = 1, \cdots, n),$$

则称矩阵序列 $\{A_k\}$ 有极限 $A = (\alpha_{ij}) \in \mathbf{C}^{m\times n}$，或称 A_k 收敛到 A，记作

$$\lim_{k\to\infty} A_k = A.$$

因此，如果把 $A \in \mathbf{C}^{m\times n}$ 看作 $\mathbf{C}^{mn}$ 中的一个向量，把 $\mathbf{C}^{m\times n}$ 上的任一矩阵范数视为 $\mathbf{C}^{mn}$ 上的一个向量范数，则据定理 1.3，有下述结论成立:

定理 2.2. 设 ν 是 $\mathbf{C}^{m\times n}$ 上的任一范数，$A, A_1, A_2, \cdots \in \mathbf{C}^{m\times n}$. 则

$$\lim_{k\to\infty} A_k = A \Longleftrightarrow \lim_{k\to\infty} \nu(A_k - A) = 0.$$

值得指出的是，范数$\| \quad \|_F$具备所谓“相容性条件”，即

$$\|AB\|_F \leqslant \|A\|_F \|B\|_F, \forall A \in \mathbf{C}^{m\times n}, \quad \forall B \in \mathbf{C}^{n\times k}.$$

事实上，如果令 $\lambda_{\max}(A^HA)$表示 A^HA 的最大特征值，则有

$$\begin{aligned}\|AB\|_F^2 &= \operatorname{tr}(B^HA^HAB) \leqslant \lambda_{\max}(A^HA)\operatorname{tr}(B^HB) \\ &\leqslant \operatorname{tr}(A^HA)\operatorname{tr}(B^HB) = \|A\|_F^2\|B\|_F^2.\end{aligned}$$

但是，范数$\| \quad \|_\infty'$和$\| \quad \|_a$都不具备“相容性条件”. 例如，对于 $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$和 $B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$，有 $\|A\|_\infty' = \|B\|_\infty' = 1$，而 $\|AB\|_\infty' = 2 > \|A\|_\infty'\|B\|_\infty'$; $\|A\|_a = \|B\|_a = \frac{3}{2}$，而

$AB = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$

$$\|AB\|_\alpha = \frac{5}{2} > \|A\|_\alpha \|B\|_\alpha.$$

在数值分析中，具备相容性条件的范数在使用上是特别方便的。因此，特地给出下列定义。

定义 2.2. 设 $\mu: \mathbb{C}^{m\times n} \to \mathbb{R}_+$, $\nu: \mathbb{C}^{n\times k} \to \mathbb{R}_+$ 和 $\rho: \mathbb{C}^{m\times k} \to \mathbb{R}^+$ 是矩阵范数。如果

$$\rho(AB) \leqslant \mu(A)\nu(B), \forall A \in \mathbb{C}^{m\times n}, \ \forall B \in \mathbb{C}^{n\times k},$$

则称 μ,ν 和 ρ 是相容的。特别地，如果 $\mathbb{C}^{n\times n}$ 上的范数 ν 满足

IV) $\qquad \nu(AB) \leqslant \nu(A)\nu(B)$ （相容性），

则称 ν 是自相容的矩阵范数，或简单地称 ν 是相容范数。

定义 2.2 包含了矩阵范数与向量范数的相容性定义。满足这种相容性的有

$$\|Ax\|_2 \leqslant \|A\|_F \|x\|_2, \ \forall A \in \mathbb{C}^{m\times n}, \ \forall x \in \mathbb{C}^n.$$

事实上，

$$\|Ax\|_2^2 = x^H A^H A x \leqslant \lambda_{\max}(A^H A)\|x\|_2^2 \leqslant \mathrm{tr}(A^H A)\|x\|_2^2$$
$$= \|A\|_F^2 \|x\|_2^2.$$

下述定理 2.3，将说明对于任一相容的矩阵范数，必存在与之相容的向量范数。

定理 2.3. 设 $\|\ \ \|: \mathbb{C}^{n\times n} \to \mathbb{R}_+$ 是一相容的矩阵范数。则在 $\mathbb{C}^n$ 上必存在与 $\|\ \ \|$ 相容的向量范数 ν。

证明:

任意取定一非零向量 $\alpha \in \mathbb{C}^n$。定义

$$\nu(x) = \|x\alpha^T\|, \ x \in \mathbb{C}^n.$$

容易验证 ν 是 $\mathbb{C}^n$ 上的向量范数，并且由

$$\nu(Ax) = \|Ax\alpha^T\| \leqslant \|A\|\|x\alpha^T\| = \|A\|\nu(x)$$

立即得知 ν 与 $\|\ \ \|$ 相容。 □

相容的矩阵范数还有一个重要的性质，即

定理 2.4. 如果 $\|\ \ \|: \mathbb{C}^{n\times n} \to \mathbb{R}_+$ 是一相容的矩阵范数，则对任一 $A \in \mathbb{C}^{n\times n}$，有

$$|\lambda_i| \leqslant \|A\|, \ \forall \lambda_i \in \lambda(A).$$

证明:

据定理 2.3, 在 $\mathbb{C}^n$ 上存在与 $\|\ \|$ 相容的向量范数 ν. 设 $\boldsymbol{x}$ 是 A 属于 λ_i 的特征向量, 即

$$A\boldsymbol{x} = \lambda_i \boldsymbol{x},\ \ \boldsymbol{x} \in \mathbb{C}^n,\ \ \boldsymbol{x} \neq 0,$$

则由

$$|\lambda_i|\nu(\boldsymbol{x}) = \nu(\lambda_i \boldsymbol{x}) = \nu(A\boldsymbol{x}) \leqslant \|A\|\nu(\boldsymbol{x})$$

和 $\nu(\boldsymbol{x}) > 0$ 立即推出定理之结论.□

从不同的角度, 可以把范数加以分类. 在数值分析中常常使用的是算子范数和酉不变范数. 下一段 (即 2.2) 讨论算子范数, § 3 讨论酉不变范数.

2.2 算子范数

首先证明两条引理.

引理 2.3. 设 ν 是 $\mathbb{C}^n$ 上的范数. 则点集

$$\mathscr{S}_\nu = \{\boldsymbol{x} \in \mathbb{C}^n : \nu(\boldsymbol{x}) = 1\}$$

是 $\mathbb{C}^n$ 内一有界闭集.

证明:

1) 据定理 1.2, 存在常数 $r_1, r_2 > 0$, 使得

$$r_1 \nu(\boldsymbol{x}) \leqslant \|\boldsymbol{x}\|_2 \leqslant r_2 \nu(\boldsymbol{x}),\ \ \forall \boldsymbol{x} \in \mathbb{C}^n. \tag{2.1}$$

由此立即得出

$$\|\boldsymbol{x}\|_2 \leqslant r_2,\ \ \forall \boldsymbol{x} \in \mathscr{S}_\nu.$$

所以 $\mathscr{S}_\nu$ 在 $\mathbb{C}^n$ 内有界.

2) 由(2.1)和

$$\nu(\boldsymbol{x}_k) - \nu(\boldsymbol{x}_k - \boldsymbol{x}) \leqslant \nu(\boldsymbol{x}) \leqslant \nu(\boldsymbol{x}_k) + \nu(\boldsymbol{x} - \boldsymbol{x}_k)$$

可知, 如果 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots$ 是 $\mathscr{S}_\nu$ 中的点列, 并且 $\|\boldsymbol{x}_k - \boldsymbol{x}\|_2 \to 0$ $(k \to \infty)$, 则有

$$1 - \frac{1}{r_1}\|\boldsymbol{x}_k - \boldsymbol{x}\|_2 \leqslant \nu(\boldsymbol{x}) \leqslant 1 + \frac{1}{r_1}\|\boldsymbol{x} - \boldsymbol{x}_k\|_2.$$

令 $k \to \infty$, 得到 $\nu(\boldsymbol{x}) = 1$, 即 $\boldsymbol{x} \in \mathscr{S}_\nu$. 所以 $\mathscr{S}_\nu$ 是闭集. □

参考证明引理 1.1 的办法，可证（作为本节的习题 1）

引理 2.4. 设 μ 是 $\mathbf{C}^m$ 上的范数，$A\in\mathbf{C}^{m\times n}$，则 $\mu(A\boldsymbol{x})$ 是 $\boldsymbol{x}\in\mathbf{C}^n$ 的连续函数。

因此，如果任取 $\mathbf{C}^n$ 上的范数 ν，$\mu(A\boldsymbol{x})$ 必在 $\mathscr{S}_\nu$ 上达到极大值。所以可以定义

$$\|A\|_{\mu,\nu}\triangleq\max_{\nu(\boldsymbol{x})=1}\mu(A\boldsymbol{x}),\quad A\in\mathbf{C}^{m\times n}. \tag{2.2}$$

定理 2.5. 设 $\mu:\mathbf{C}^m\to\mathbf{R}_+$ 和 $\nu:\mathbf{C}^n\to\mathbf{R}_+$ 是向量范数。如果按 (2.2) 定义 $\|\ \ \|_{\mu,\nu}:\mathbf{C}^{m\times n}\to\mathbf{R}_+$，则 $\|\ \ \|_{\mu,\nu}$ 是 $\mathbf{C}^{m\times n}$ 上的范数。

证明：

首先由 (2.2) 可知

$$\|A\|_{\mu,\nu}=\sup_{\substack{\boldsymbol{x}\in\mathbf{C}^n\\ \boldsymbol{x}\neq 0}}\frac{\mu(A\boldsymbol{x})}{\nu(\boldsymbol{x})}.$$

因此

$$\mu(A\boldsymbol{x})\leqslant\|A\|_{\mu,\nu}\nu(\boldsymbol{x}),\forall A\in\mathbf{C}^{m\times n},\forall\boldsymbol{x}\in\mathbf{C}^n. \tag{2.3}$$

以下证明：$\|\ \ \|_{\mu,\nu}$ 满足矩阵范数定义中的性质 I)，II) 和 III)。

1) 正定性. 设 $A\neq 0$，则必有 i，使 $A\boldsymbol{e}_i\neq 0$，此处 $\boldsymbol{e}_i$ 表示 $I^{(n)}$ 的第 i 列向量。于是由 (2.3) 知

$$0<\mu(A\boldsymbol{e}_i)\leqslant\|A\|_{\mu,\nu}\nu(\boldsymbol{e}_i)\Longrightarrow\|A\|_{\mu,\nu}>0.$$

2) 齐次性. 任取 $\alpha\in\mathbf{C}$，有

$$\|\alpha A\|_{\mu,\nu}=\max_{\nu(\boldsymbol{x})=1}\mu(\alpha A\boldsymbol{x})=\max_{\nu(\boldsymbol{x})=1}|\alpha|\mu(A\boldsymbol{x})$$
$$=|\alpha|\,\|A\|_{\mu,\nu}.$$

3) 三角不等式. 任取 $A,B\in\mathbf{C}^{m\times n}$. 设 $\boldsymbol{x}$ 满足 $\nu(\boldsymbol{x})=1$ 并且 $\mu((A+B)\boldsymbol{x})=\|A+B\|_{\mu,\nu}$，则由 (2.3) 知

$$\|A+B\|_{\mu,\nu}=\mu((A+B)\boldsymbol{x})\leqslant\mu(A\boldsymbol{x})+\mu(B\boldsymbol{x})$$
$$\leqslant\|A\|_{\mu,\nu}\nu(\boldsymbol{x})+\|B\|_{\mu,\nu}\nu(\boldsymbol{x})=\|A\|_{\mu,\nu}+\|B\|_{\mu,\nu}. \quad\square$$

根据定理 2.5，可以引进下述定义。

定义 2.3. 设 μ 与 ν 分别是 $\mathbf{C}^m$ 与 $\mathbf{C}^n$ 上的范数，则由 (2.2) 所定义的 $\|\ \ \|_{\mu,\nu}:\mathbf{C}^{m\times n}\to\mathbf{R}_+$ 叫做 $\mathbf{C}^{m\times n}$ 上的算子范数，或从属于 μ 与 ν 的范数。

关于算子范数的相容性，有下述结果.

定理 2.6. 设 μ,ν 与 ω 分别是 $\mathbb{C}^m$, $\mathbb{C}^n$ 与 $\mathbb{C}^k$ 上的范数. 如果按照 (2.2) 分别定义 $\mathbb{C}^{m\times n}$, $\mathbb{C}^{n\times k}$ 与 $\mathbb{C}^{m\times k}$ 上的算子范数 $\|\ \|_{\mu,\nu}$, $\|\ \|_{\nu,\omega}$ 与 $\|\ \|_{\mu,\omega}$, 则有

$$\|AB\|_{\mu,\omega} \leqslant \|A\|_{\mu,\nu}\|B\|_{\nu,\omega},\ \forall A\in \mathbb{C}^{m\times n},\ \forall B\in \mathbb{C}^{n\times k}.$$

证明:

设 $\boldsymbol{u}\in \mathbb{C}^k$, $\omega(\boldsymbol{u})=1$ 并且 $\mu(AB\boldsymbol{u})=\|AB\|_{\mu,\omega}$, 则由不等式(2.3)得出

$$\begin{aligned}\|AB\|_{\mu,\omega} &= \mu(AB\boldsymbol{u}) = \mu(A(B\boldsymbol{u})) \leqslant \|A\|_{\mu,\nu}\nu(B\boldsymbol{u})\\ &\leqslant \|A\|_{\mu,\nu}\|B\|_{\nu,\omega}\omega(\boldsymbol{u}) = \|A\|_{\mu,\nu}\|B\|_{\nu,\omega}.\ \square\end{aligned}$$

当 $m=n=k$ 时，如果 μ,ν 和 ω 为 $\mathbb{C}^n$ 上的同一个范数，则从定理 2.6 立即得到

定理 2.7. 设 ν 是 $\mathbb{C}^n$ 上的范数, 则在 $\mathbb{C}^{n\times n}$ 上从属于 ν 的矩阵范数 $\|\ \|_\nu$ 是相容范数，即

$$\|AB\|_\nu \leqslant \|A\|_\nu\|B\|_\nu,\ \forall A,\ B\in \mathbb{C}^{n\times n}.$$

此外，还有

定理 2.8. 设 ν 是 $\mathbb{C}^n$ 上的范数，$\|\ \|_\nu$ 是 $\mathbb{C}^{n\times n}$ 上从属于 ν 的矩阵范数. 又设 $\|\ \|$ 是 $\mathbb{C}^{n\times n}$ 上与 ν 相容的任一种矩阵范数. 则 $\|A\|_\nu \leqslant \|A\|$, $\forall A\in \mathbb{C}^{n\times n}$.

证明:

设 $\boldsymbol{x}\in \mathbb{C}^n$ 满足 $\nu(\boldsymbol{x})=1$, 并且 $\nu(A\boldsymbol{x})=\|A\|_\nu$. 则

$$\|A\|_\nu = \nu(A\boldsymbol{x}) \leqslant \|A\|\nu(\boldsymbol{x}) = \|A\|.\ \square$$

注意到，如果把 $\mathbb{C}^n$ 上的范数 $\|\ \|_1$, $\|\ \|_2$ 与 $\|\ \|_\infty$ 限制到 $\mathbb{C}^l(l<n)$ 上，恰好是 $\mathbb{C}^l$ 上的范数 $\|\ \|_1$, $\|\ \|_2$ 与 $\|\ \|_\infty$，所以，根据定理 2.5，可以得到 $\mathbb{C}^{m\times n}$ 上的算子范数 $\|\ \|_p$:

$$\|A\|_p = \max_{\|\boldsymbol{x}\|_p=1}\|A\boldsymbol{x}\|_p,\ A\in \mathbb{C}^{m\times n},\ p=1,2,\infty; \tag{2.4}$$

而且由定理 2.6 可知，这些算子范数都是相容范数，即

$$\|AB\|_p \leqslant \|A\|_p\|B\|_p,\ A\in \mathbb{C}^{m\times n},\ B\in \mathbb{C}^{n\times k},\ p=1,2,\infty. \tag{2.5}$$

关于 $\|A\|_p$，有下面的表示定理。

定理 2.9. 设 $A=(\alpha_{ij})\in \mathbf{C}^{m\times n}$，则有

$$\|A\|_1=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|\alpha_{ij}|, \tag{2.6}$$

$$\|A\|_\infty=\max_{1\leqslant i\leqslant m}\sum_{j=1}^{n}|\alpha_{ij}|. \tag{2.7}$$

和

$$\|A\|_2=\sqrt{\lambda_{\max}(A^HA)}. \tag{2.8}$$

证明:

记 $A=(\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_n)$，其中 $\boldsymbol{\alpha}_j\in\mathbf{C}^m$，$j=1,\cdots,n$。任取 $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T\neq 0$，有

$$\|A\boldsymbol{x}\|_1=\left\|\sum_{i=1}^{n}\xi_i\boldsymbol{\alpha}_i\right\|_1\leqslant\sum_{i=1}^{n}|\xi_i|\,\|\boldsymbol{\alpha}_i\|_1\leqslant\max_{1\leqslant j\leqslant n}\|\boldsymbol{\alpha}_j\|_1\cdot\|\boldsymbol{x}\|_1.$$

上式表示

$$\|A\|_1\leqslant\max_{1\leqslant j\leqslant n}\|\boldsymbol{\alpha}_j\|_1. \tag{2.9}$$

另一方面，如果 $\max\limits_{1\leqslant j\leqslant n}\|\boldsymbol{\alpha}_j\|_1=\|\boldsymbol{\alpha}_k\|_1$，则由 $\|\boldsymbol{e}_k\|_1=1$（$\boldsymbol{e}_k$ 表示 $I^{(n)}$ 的第 k 列向量）和

$$\frac{\|A\boldsymbol{e}_k\|_1}{\|\boldsymbol{e}_k\|_1}=\|\boldsymbol{\alpha}_k\|_1=\max_{1\leqslant j\leqslant n}\|\boldsymbol{\alpha}_j\|_1$$

知

$$\|A\|_1\geqslant\max_{1\leqslant j\leqslant n}\|\boldsymbol{\alpha}_j\|_1. \tag{2.10}$$

把(2.9)与(2.10)联系起来，便证明了(2.6)。

同理可证(2.7)(作为本节习题 2)。

关于(2.8)的证明，只需利用 A 的奇异值分解：

$$A=U\begin{pmatrix}\Sigma_1 & 0\\ 0 & 0\end{pmatrix}V^H,\ \Sigma_1=\operatorname{diag}(\sigma_1,\cdots,\sigma_r)>0,$$

其中 U 与 V 为酉阵。按照定义(2.4)，有

$$\|A\|_2^2 = \max_{\|x\|_2=1} x^H A^H A x = \max_{\|x\|_2=1} (V^H x)^H \begin{pmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix} V^H x$$

$$= \max_{\|y\|_2=1} y^H \begin{pmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{pmatrix} y = \lambda_{\max}(A^H A). \quad \square$$

鉴于$\|A\|_p(p=1,2,\infty)$的表示式(2.6)—(2.8)，通常把$\|A\|_1$和$\|A\|_\infty$分别叫做A的列和范数和行和范数，把$\|A\|_2$叫做A的谱范数.

矩阵的谱范数不便计算，但它有许多很好的性质，在理论研究中是常常用到的. 下述定理列举了它的若干性质.

定理 2.10. 设 $A \in \mathbf{C}^{m\times n}$. 则

$$\|A\|_2 = \max_{\substack{\|x\|_2=1 \\ \|y\|_2=1}} |y^H A x|, \tag{2.11}$$

$$\|A^H\|_2 = \|A^T\|_2 = \|A\|_2, \tag{2.12}$$

$$\|A^H A\|_2 = \|A\|_2^2, \tag{2.13}$$

$$\|A\|_2^2 \leqslant \|A\|_1 \|A\|_\infty, \tag{2.14}$$

并且对于任意的酉阵 $U \in \mathbf{C}^{m\times m}$ 和 $V \in \mathbf{C}^{n\times n}$，有

$$\|UAV\|_2 = \|A\|_2. \tag{2.15}$$

证明:

1) 设 $x \in \mathbf{C}^n$ 与 $y \in \mathbf{C}^m$ 满足 $\|x\|_2 = \|y\|_2 = 1$，则由 Cauchy 不等式，有

$$|y^H A x| \leqslant \|y\|_2 \|Ax\|_2 \leqslant \|A\|_2 \|x\|_2 = \|A\|_2;$$

另一方面，设 $\|x\|_2 = 1$ 并且 $\|Ax\|_2 = \|A\|_2$，则若取

$$y = \frac{Ax}{\|Ax\|_2},$$

便有

$$|y^H A x| = \frac{x^H A^H A x}{\|Ax\|_2} = \|Ax\|_2 = \|A\|_2.$$

因此(2.11)式成立.

2) 利用(2.11)，可得

$$\|A^H\|_2 = \max_{\|x\|_2=\|y\|_2=1} |y^H A^H x| = \max_{\|x\|_2=\|y\|_2=1} |x^H A y|$$

$$= \|A\|_2.$$

因此(2.12)成立.

3) 利用(2.5)和(2.12),

$$\|A^HA\|_2 \leqslant \|A^H\|_2\|A\|_2 = \|A\|_2^2;$$

另一方面,取 $\boldsymbol{x}$ 满足 $\|\boldsymbol{x}\|_2 = 1$ 和 $\|A\boldsymbol{x}\|_2 = \|A\|_2$, 并在(2.11)中取 $\boldsymbol{y} = \boldsymbol{x}$,则

$$\|A^HA\|_2 \geqslant \boldsymbol{x}^HA^HA\boldsymbol{x} = \|A\boldsymbol{x}\|_2^2 = \|A\|_2^2.$$

因此(2.13)成立.

4) 利用(2.8),有

$$\|A\|_2^2 = \lambda_{\max}(A^HA);$$

再根据(2.5), $\|\ \ \|_1$ 是相容的矩阵范数,因此由定理 2.4 可得

$$\lambda_{\max}(A^HA) \leqslant \|A^HA\|_1 \leqslant \|A^H\|_1\|A\|_1 = \|A\|_\infty\|A\|_1.$$

所以(2.14)成立.

5) 由(2.8)和

$$\lambda_{\max}[(UAV)^H(UAV)] = \lambda_{\max}(A^HA)$$

立即得到等式(2.15). □

定义 2.4. 设 $A \in \mathbf{C}^{n\times n}$. 则

$$\rho(A) = \max\{|\lambda| : \lambda \in \lambda(A)\}$$

叫做 A 的谱半径.

定理 2.4 表明,如果 $\|\ \ \|$ 是 $\mathbf{C}^{n\times n}$ 上的相容范数,则

$$\rho(A) \leqslant \|A\|, \ \forall A \in \mathbf{C}^{n\times n}. \tag{2.16}$$

另一方面,有下述重要结果:

定理 2.11. 设 $A \in \mathbf{C}^{n\times n}$, $\varepsilon > 0$. 则存在 $\mathbf{C}^{n\times n}$ 上的相容范数 $\|\ \ \|$(依赖于 A 和 ε), 满足 $\|I^{(n)}\| = 1$, 使得

$$\|A\| \leqslant \rho(A) + \varepsilon. \tag{2.17}$$

证明:

根据第一章定理 1.2 (Schur 定理), 存在上三角阵 $T = U^HAU$, 其中 U 为酉阵, $T = \Lambda + M$, $\Lambda = \mathrm{diag}\,(\lambda_1, \cdots, \lambda_n)$, $\lambda_1, \cdots, \lambda_n$ 为 A 的特征值, $M = (m_{ij})$ 为严格上三角阵. 取

$$\delta = \min\left\{1, \frac{\varepsilon}{(n-1)\max\limits_{1\leqslant i<j\leqslant n}|m_{ij}|}\right\},$$

令 $D = \text{diag}\,(1,\delta,\delta^2,\cdots,\delta^{n-1})$，则

$$D^{-1}TD = \begin{pmatrix} \lambda_1 & m_{12}\delta & m_{13}\delta^2\cdots m_{1n}\delta^{n-1} \\ & \lambda_2 & m_{23}\delta\cdots m_{2n}\delta^{n-2} \\ & & \ddots \quad \vdots \\ 0 & & \lambda_n \end{pmatrix}.$$

于是

$$\begin{aligned}\|D^{-1}U^HAUD\|_\infty &= \|D^{-1}TD\|_\infty \\ &\leqslant \max_i|\lambda_i| + \max_{i<j}|m_{ij}|(1+\delta+\cdots+\delta^{n-2})\delta \\ &\leqslant \rho(A) + (n-1)\max_{i<j}|m_{ij}|\delta \leqslant \rho(A)+\varepsilon. \qquad (2.18)\end{aligned}$$

现定义 $\mathbf{C}^{n\times n}$ 到 $\mathbf{R}_+$ 的函数$\|\ \ \|$:

$$\|G\| = \|D^{-1}U^HGUD\|_\infty,\ \forall G\in\mathbf{C}^{n\times n}.$$

易知$\|\ \ \|$是 $\mathbf{C}^{n\times n}$ 上的相容范数，满足 $\|I^{(n)}\|=1$，并且由(2.18)知(2.17)式成立. □

注 2.1. 与(2.18)式类似，下式成立:

$$\|D^{-1}U^HAUD\|_1 \leqslant \rho(A)+\varepsilon. \qquad (2.19)$$

因此，也可以定义 $\mathbf{C}^{n\times n}$ 到 $\mathbf{R}_+$ 的函数$\|\ \ \|$:

$$\|G\| = \|D^{-1}U^HGUD\|_1,\ \forall G\in\mathbf{C}^{n\times n},$$

$\|\ \ \|$ 也是 $\mathbf{C}^{n\times n}$ 上的相容范数，满足 $\|I\|=1$，并且由(2.19)可得(2.17).

利用定理 2.2 和定理 2.11，可证下述结论(作为本节习题 3):

定理 2.12. 设 $A\in\mathbf{C}^{n\times n}$. 则

$$\lim_{k\to\infty}A^k = 0 \Longleftrightarrow \rho(A)<1.$$

进而可得

定理 2.13. 设 $A\in\mathbf{C}^{n\times n}$. 则

$$\sum_{k=0}^{\infty}A^k\ 收敛 \Longleftrightarrow \lim_{k\to\infty}A^k=0; \qquad (2.20)$$

而且当(2.20)右端条件被满足时，有

$$\sum_{k=0}^{\infty} A^k = (I - A)^{-1}. \tag{2.21}$$

此外，对于收敛级数 $\sum_{k=0}^{\infty} A^k$ 的部分和，有下述估计，即：存在 $\mathbf{C}^{n\times n}$ 上的相容范数$\|\ \ \|$，使得

$$\|(I - A)^{-1} - \sum_{k=0}^{m} A^k\| \leqslant \frac{\|A\|^{m+1}}{1 - \|A\|}. \tag{2.22}$$

证明:

"$\Longrightarrow$"

记 $A^k = (\alpha_{ij}^{(k)})$, $k = 0, 1, 2, \cdots$. $\sum_{k=0}^{\infty} A^k$ 收敛，即指每个级数 $\sum_{k=0}^{\infty} \alpha_{ij}^{(k)}$ 都收敛 $(1 \leqslant i, j \leqslant n)$. 因此，$\lim_{k\to\infty} \alpha_{ij}^{(k)} = 0$ $(1 \leqslant i, j \leqslant n)$，即 $\lim_{k\to\infty} A^k = 0$.

"$\Longleftarrow$"

根据定理 2.12，由 $\lim_{k\to\infty} A^k = 0$ 知 $\rho(A) < 1$，因而 $I - A$ 为非奇异阵. 于是由

$$\sum_{k=0}^{m} A^k(I - A) = I - A^{m+1}$$

得

$$\sum_{k=0}^{m} A^k = (I - A)^{-1} - A^{m+1}(I - A)^{-1}. \tag{2.23}$$

再在上式两端取极限 $k \to \infty$，即得到 (2.21). 这表示级数 $\sum_{k=0}^{\infty} A^k$ 收敛.

不等式(2.22)可如下证明. 据 $\rho(A) < 1$ 和定理 2.11，存在 $\mathbf{C}^{n\times n}$ 上的相容范数$\|\ \ \|$，满足 $\|I\| = 1$，并使得 $\|A\| < 1$. 将关系式(2.23)改写为

$$(I - A)^{-1} - \sum_{k=0}^{m} A^k = A^{m+1}(I - A)^{-1},$$

从而

$$\left\|(I-A)^{-1}-\sum_{k=0}^{m} A^k\right\| \leqslant \|A^{m+1}\|\|(I-A)^{-1}\|$$

$$\leqslant \|A\|^{m+1}\|(I-A)^{-1}\|. \quad (2.24)$$

再由

$$(I-A)^{-1}=I+A(I-A)^{-1}$$

知

$$\|(I-A)^{-1}\| \leqslant \|I\|+\|A(I-A)^{-1}\|$$

$$\leqslant 1+\|A\|\|(I-A)^{-1}\|,$$

即

$$\|(I-A)^{-1}\| \leqslant (1-\|A\|)^{-1}. \quad (2.25)$$

将(2.25)代入(2.24),便得出(2.22). □

习题

1. 设 μ 是 $\mathbf{C}^m$ 上的范数，$A \in \mathbf{C}^{m\times n}$. 试证 $\mu(A\boldsymbol{x})$ 是 $\boldsymbol{x} \in \mathbf{C}^n$ 的连续函数.

2. 设 $A=(\alpha_{ij}) \in \mathbf{C}^{m\times n}$, 则 $\|A\|_\infty = \max\limits_{1\leqslant i\leqslant m} \sum\limits_{j=1}^{n} |\alpha_{ij}|$.

3. 设 $A \in \mathbf{C}^{n\times n}$. 则

$$\lim_{k\to\infty} A^k = 0 \Longleftrightarrow \rho(A) < 1.$$

4. 设 ν 是 $\mathbf{C}^n$ 上的范数，$B \in \mathbf{C}^{n\times n}$ 是非奇异阵，μ 表示 $\mathbf{C}^n$ 上由 $\mu(\boldsymbol{x})=\nu(B\boldsymbol{x})$定义的范数. 令$\|\ \|_\nu$ 和$\|\ \|_\mu$ 分别表示 $\mathbf{C}^{n\times n}$ 上从属于 ν 和 μ 的算子范数. 试证

$$\|A\|_\mu = \|BAB^{-1}\|_\nu, \ \forall A \in \mathbf{C}^{n\times n}.$$

5. 证明: $\rho(A) < 1 \Longleftrightarrow$ 存在正定阵 Q，使得 $Q - AQA^H$ 为正定阵.

6. 特征值的实部均为负数的矩阵叫做稳定矩阵. 试利用第 5 题的结论证明: $-B$ 为稳定矩阵的充分必要条件是存在正定阵 M,使得 $BM + MB^H > 0$.

7. 设 A 为 Hermite 阵. 则

$$\|A\|_2 \leqslant \|A\|_\infty = \|A\|_1 \leqslant n^{\frac{1}{2}}\|A\|_2.$$

8. 设 $A \in \mathbf{C}^{n\times n}$, $\lambda(A) = \{\lambda_i\}$. 试证

$$\inf_{S\in \mathbf{C}_n^{n\times n}} \|S^{-1}AS\|_F^2 = \sum_{i=1}^{n} |\lambda_i|^2,$$

并且上式中的下确界可以达到的充分必要条件是 A 可以正规化.

9. 设 P 是幂等阵(即 $P^2 = P$),则对任一相容的矩阵范数$\|\ \|$, 必有 $\|P\| \geqslant 1$; 此外, 若幂等阵 $P \neq I$, 则必有相容的矩阵范数 $\|\ \|$, 使得 $\|P\| > 1$.

10. 设函数$\|\|$: $\mathbf{C}^{n\times n} \to \mathbf{R}_+$, 其定义为

$$\|A\| = n\max_{i,j}|\alpha_{ij}|, \ \forall A = (\alpha_{ij}) \in \mathbf{C}^{n\times n}.$$

试证: $\|\ \|$ 是 $\mathbf{C}^{n\times n}$ 上的相容范数.

11. 设 ν 是 $\mathbf{C}^{n\times n}$ 上的范数. 试证: 存在 $\sigma > 0$, 使得由

$$\|A\| = \sigma\nu(A)$$

所定义的$\|\ \|$是 $\mathbf{C}^{n\times n}$ 上的相容范数.

12. 设 $B \in \mathbf{C}^{n\times n}$, $\boldsymbol{x}_0$ 与 $\boldsymbol{d} \in \mathbf{C}^n$, $\|\ \|$ 是 $\mathbf{C}^{n\times n}$ 上的相容范数, 满足 $\|I\| = 1$. 构造序列 $\{\boldsymbol{x}_k\}$ 如下:

$$\boldsymbol{x}_{k+1} = B\boldsymbol{x}_k + \boldsymbol{d}, k = 0, 1, 2, \cdots. \tag{i}$$

试证: 如果 $\|B\| < 1$, 则由 (i) 式构造的序列 $\{\boldsymbol{x}_k\}$ 收敛到 $\hat{\boldsymbol{x}} \in \mathbf{C}^n$,并且对于 $\mathbf{C}^n$ 上与$\|\ \|$相容的向量范数 ν, 有

$$\nu(\hat{\boldsymbol{x}} - \boldsymbol{x}_k) \leqslant \frac{\|B\|^k}{1 - \|B\|}(\nu(\boldsymbol{d}) + \nu(\boldsymbol{x}_0)).$$

§3 $\mathbb{C}^{m\times n}$ 上的酉不变范数

3.1 定义

以下用 $\mathscr{U}_n$ 表示全体 $n \times n$ 酉阵的集合.

定义 3.1. 一个定义在 $\mathbf{C}^{m\times n}$ 上的非负实值函数$\|\ \|$叫做 $\mathbf{C}^{m\times n}$ 上的酉不变范数,如果它满足下列条件

I) $A \neq 0 \Longrightarrow \|A\| > 0$,

II) $\|\alpha A\| = |\alpha| \, \|A\|, \forall \alpha \in \mathbf{C}$,

III) $\|A + B\| \leqslant \|A\| + \|B\|$,

IV) $\|UAV\| = \|A\|$, $\forall U \in \mathscr{U}_m$, $\forall V \in \mathscr{U}_n$,

V) $\|A\| = \|A\|_2, \forall A$ 满足 $\operatorname{rank}(A) = 1$.

I)—V) 中的 A, B 为 $\mathbf{C}^{m\times n}$ 中的任意矩阵.

注 3.1. 定义 3.1 中的条件 V),等价于下述条件

V)′ 若 $A = \boldsymbol{x}\boldsymbol{y}^H, \boldsymbol{x} \in \mathbf{C}^m$, $\boldsymbol{y} \in \mathbf{C}^n$, $\boldsymbol{x} \neq 0, \boldsymbol{y} \neq 0$, 则

$$\|A\| = \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2.$$

其等价性证明如下:

V)′ $\Longrightarrow$ V). 若 $\operatorname{rank} A = 1$, 则由第一章定理 3.1, $A = \boldsymbol{x}\boldsymbol{y}^H$, 其中 $\boldsymbol{x} \in \mathbf{C}^m$, $\boldsymbol{y} \in \mathbf{C}^n$, 并且 $\boldsymbol{x} \neq 0$, $\boldsymbol{y} \neq 0$. 据 (2.8),

$$\begin{aligned}\|A\|_2^2 &= \lambda_{\max}(\boldsymbol{y}\boldsymbol{x}^H\boldsymbol{x}\boldsymbol{y}^H) = \|\boldsymbol{x}\|_2^2 \lambda_{\max}(\boldsymbol{y}\boldsymbol{y}^H) \\ &= \|\boldsymbol{x}\|_2^2 \|\boldsymbol{y}\|_2^2 \lambda_{\max}(\boldsymbol{u}_1\boldsymbol{u}_1^H),\end{aligned}$$

其中 $\boldsymbol{u}_1 = (\boldsymbol{y}^H\boldsymbol{y})^{-\frac{1}{2}}\boldsymbol{y}$ 为单位向量. 易知 $\lambda_{\max}(\boldsymbol{u}_1\boldsymbol{u}_1^H) = 1$, 因而

$$\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 = \|A\|_2. \tag{3.1}$$

于是由 V)′ 导出了 V).

V) $\Longrightarrow$ V)′. 若 $A = \boldsymbol{x}\boldsymbol{y}^H$, 并且 $\boldsymbol{x} \neq 0$, $\boldsymbol{y} \neq 0$, 则

$$\operatorname{rank}(A) = 1.$$

根据 V)及(3.1),立即推知 V)′ 成立. □

例 3.1. 矩阵的谱范数 $\| \ \|_2$ 和 Frobenius 范数 $\| \ \|_F$ 是酉不变范数. 但 $\| \ \|_1$ 与 $\| \ \|_\infty$ 不是酉不变范数.

值得注意的是,如果利用 A 的奇异值分解 $A = U\Sigma V^H$, 其中 $U \in \mathscr{U}_m, V \in \mathscr{U}_n$, $\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}$, $\Sigma_1 = \operatorname{diag}(\sigma_1, \cdots, \sigma_r) > 0$, $r \leqslant \min\{m, n\}$, 则对于 $\mathbf{C}^{m\times n}$ 上的任一酉不变范数 $\| \ \|$, 有

$$\|A\| = \|\Sigma\|.$$

特别地,如果 $A \in \mathbf{C}^{m\times n}$ 的非零奇异值为 $\sigma_1, \sigma_2, \cdots, \sigma_r$, 则

$$\|A\|_2 = \max_i \sigma_i, \quad \|A\|_F = \sqrt{\sum_{i=1}^r \sigma_i^2}.$$

可见，酉不变范数必是奇异值的函数。因此启发了一个问题：如何利用奇异值的一个函数类，去刻划矩阵的酉不变范数？ von Neumann 回答了这个问题。

下面两段(3.2 和 3.3)，是为了论证 von Neumann 的结果作准备。

3.2 von Neumann 不等式

引理 3.1. 设 $A\in\mathbb{C}^{n\times n}$. 则

$$\operatorname{Re}\operatorname{tr}(A)\geqslant\operatorname{Re}\operatorname{tr}(WA),\ \forall W\in\mathscr{U}_n\Longleftrightarrow A\geqslant 0.$$

证明:

"$\Longleftarrow$"

记 A 的酉分解式为 $A=V\Lambda V^H$，其中 $\Lambda=\operatorname{diag}(\alpha_1,\cdots,\alpha_n)\geqslant 0, V\in\mathscr{U}_n$. 设 W 是任一酉阵，令 $U=V^HWV=(\mu_{ij})\in\mathscr{U}_n$. 于是有

$$\begin{aligned}\operatorname{Re}\operatorname{tr}(WA)&=\operatorname{Re}\operatorname{tr}(WV\Lambda V^H)=\operatorname{Re}\operatorname{tr}(U\Lambda)\\&=\operatorname{Re}\sum_{i=1}^n\alpha_i\mu_{ii}\leqslant\sum_{i=1}^n\alpha_i|\mu_{ii}|\leqslant\sum_{i=1}^n\alpha_i=\operatorname{tr}(A).\end{aligned}$$

"$\Longrightarrow$"

设 $\operatorname{rank}A=r$. 于是 A 有奇异值分解 $A=U\Sigma V^H$，其中

$$U=(\underset{r}{U_1},\underset{n-r}{U_2}),\qquad V=(\underset{r}{V_1},\underset{n-r}{V_2})\in\mathscr{U}_n,\Sigma=\begin{pmatrix}\Sigma_1&0\\0&0\end{pmatrix},$$

$\Sigma_1=\operatorname{diag}(\sigma_1,\cdots,\sigma_r)>0$. 如果记 $V^HU=W_0=(\omega_{ij})$，则

$$\operatorname{Re}\operatorname{tr}A=\operatorname{Re}\operatorname{tr}(U\Sigma V^H)=\operatorname{Re}\operatorname{tr}(W_0\Sigma)=\sum_{i=1}^n\sigma_i\operatorname{Re}(\omega_{ii});$$

另一方面，特别地取 $W=VU^H$，则有

$$\operatorname{Re}\operatorname{tr}(WA)=\operatorname{Re}\operatorname{tr}(VU^HU\Sigma V^H)=\sum_{i=1}^r\sigma_i.$$

据引理假设，应有

$$\sum_{i=1}^r\sigma_i\operatorname{Re}(\omega_{ii})\geqslant\sum_{i=1}^r\sigma_i.$$

因此必须 $\mathrm{Re}(\omega_{ii}) = 1, i = 1, \cdots, r$. 但又因为 $|\omega_{ii}| \leqslant 1$,所以必有 $\omega_{ii} = 1, i = 1, \cdots, r$. 即 $V_1^H U_1 = I^{(r)}$. 从而得到 $V_1 = U_1$, $A = U_1 \Sigma_1 U_1^H \geqslant 0$. □

引理 3.2. 设 $\Lambda = \mathrm{diag}(\alpha_1, \cdots, \alpha_n) \geqslant 0$, $\Omega = \mathrm{diag}(\beta_1, \cdots, \beta_n)$,其中 $\beta_1 > \beta_2 > \cdots > \beta_n > 0$. 如果存在 $U_0, V_0 \in \mathscr{U}_n$,使得

$$\mathrm{Re}\ \mathrm{tr}(U_0 \Lambda V_0 \Omega) \geqslant \mathrm{Re}\ \mathrm{tr}(U \Lambda V \Omega),\ \forall U, V \in \mathscr{U}_n,$$

则 $U_0 \Lambda V_0$ 必为半正定实对角阵.

证明:

据题设, 若分别取 $U = WU_0$, $V = V_0$ 和 $U = U_0$, $V = V_0 W$, 其中 $W \in \mathscr{U}_n$, 则有

$$\mathrm{Re}\ \mathrm{tr}(U_0 \Lambda V_0 \Omega) \geqslant \mathrm{Re}\ \mathrm{tr}(WU_0 \Lambda V_0 \Omega),\ \forall W \in \mathscr{U}_n$$

和

$$\mathrm{Re}\ \mathrm{tr}(\Omega U_0 \Lambda V_0) \geqslant \mathrm{Re}\ \mathrm{tr}(W \Omega U_0 \Lambda V_0),\ \forall W \in \mathscr{U}_n.$$

利用引理 3.1, $U_0 \Lambda V_0 \Omega$ 与 $\Omega U_0 \Lambda V_0$ 皆为半正定阵. 令 $U_0 \Lambda V_0 = \Gamma$,便有

$$\Gamma \Omega = \Omega \Gamma^H \geqslant 0,\ \Omega \Gamma = \Gamma^H \Omega \geqslant 0.$$

由此可得

$$\Gamma \Omega^2 = \Omega \Gamma^H \Omega = \Omega^2 \Gamma,$$

因而 $\Gamma = \mathrm{diag}(\gamma_1, \cdots, \gamma_n)$. 再由 $\Gamma \Omega \geqslant 0$ 和 $\Omega > 0$,便知 $\gamma_1, \cdots, \gamma_n \geqslant 0$. □

引理 3.3. 设 $\alpha_1 \geqslant \cdots \geqslant \alpha_n$, $\beta_1 \geqslant \cdots \geqslant \beta_n$. 若 $\pi(1), \cdots, \pi(n)$ 是 $1, \cdots, n$ 的任一排列,则

$$\sum_{i=1}^{n} \alpha_{\pi(i)} \beta_i \leqslant \sum_{i=1}^{n} \alpha_i \beta_i. \tag{3.2}$$

证明:

假设对于 $i = 1, \cdots, r$,已有 $\pi(i) = i$,但 $\pi(r+1) = r + t$ $(t > 1)$, 而 $\pi(r+s) = r + 1 (s > 1)$. 现将 $\sum_i \alpha_{\pi(i)} \beta_i$ 中的 $\alpha_{\pi(r+1)} \beta_{r+1}$ 和 $\alpha_{\pi(r+s)} \beta_{r+s}$ 分别换成 $\alpha_{\pi(r+s)} \beta_{r+1}$ 和 $\alpha_{\pi(r+1)} \beta_{r+s}$, 其余各项不变. 由

$$
\begin{aligned}
&(\alpha_{\pi(r+s)}\beta_{r+1} + \alpha_{\pi(r+1)}\beta_{r+s}) - (\alpha_{\pi(r+1)}\beta_{r+1} + \alpha_{\pi(r+s)}\beta_{r+s}) \\
&\quad = \alpha_{r+1}\beta_{r+1} + \alpha_{r+t}\beta_{r+s} - \alpha_{r+t}\beta_{r+1} - \alpha_{r+1}\beta_{r+s} \\
&\quad = (\alpha_{r+1} - \alpha_{r+t})(\beta_{r+1} - \beta_{r+s}) \geqslant 0
\end{aligned}
$$

知

$$
\sum_{i=r+1}^{n} \alpha_{\pi(i)}\beta_i \leqslant \alpha_{r+1}\beta_{r+1} + \sum_{i=r+2}^{n} \alpha_{\pi'(i)}\beta_i,
$$

其中 $\pi'(r+2),\cdots,\pi'(n)$ 是 $r+2,\cdots,n$ 的一个排列．依次类推，即可得到(3.2)．□

定理 3.1 (von Neumann [128])．设 $A, B \in \mathbf{C}^{m\times n}$，它们分别有奇异值 $\alpha_1 \geqslant \cdots \geqslant \alpha_n \geqslant 0$ 和 $\beta_1 \geqslant \cdots \geqslant \beta_n \geqslant 0$．则

$$
\max_{\substack{U\in \mathscr{U}_m \\ V\in \mathscr{U}_n}} \mathrm{Re}\,\mathrm{tr}(UAVB^H) = \sum_{i=1}^{n} \alpha_i\beta_i. \tag{3.3}
$$

证明：

以下仅证明当 $m = n$ 时，(3.3)成立．当 $m \neq n$ 时，可先将 A 与 B 补以若干行列的零元素，构成方阵，然后引用 $m = n$ 时的结论证明之（见本节习题 1）．

将 A 与 B 的奇异值分解代入(3.3)左端，并无妨假设 $A = \Lambda = \mathrm{diag}(\alpha_1,\cdots,\alpha_n) \geqslant 0$，$B = \Omega = \mathrm{diag}(\beta_1,\cdots,\beta_n) \geqslant 0$．

首先考虑

$$
\alpha_1 \geqslant \cdots \geqslant \alpha_n \geqslant 0,\ \beta_1 > \cdots > \beta_n > 0 \tag{3.4}
$$

的情形．

注意到 $\mathscr{U}_n$ 是 $\mathbf{C}^{n^2}$ 空间中的有界闭集，而 $\mathrm{Re}\,\mathrm{tr}(U\Lambda V\Omega)$是 U 和 V 的连续函数，因此它在直积空间 $\mathscr{U}_n \otimes \mathscr{U}_n$ 上可以达到极大值．设其极大值在 U_0 与 V_0 达到，据引理 3.2，有

$$
U_0\Lambda V_0 = \Gamma = \mathrm{diag}(\gamma_1,\cdots,\gamma_n) \geqslant 0. \tag{3.5}
$$

于是

$$
\max_{U,V\in \mathscr{U}_n} \mathrm{Re}\,\mathrm{tr}(U\Lambda V\Omega) = \mathrm{tr}(\Gamma\Omega) = \sum_{i=1}^{n} \gamma_i\beta_i. \tag{3.6}
$$

再由(3.5)知

$$U_0\Lambda^2U_0^H = \Gamma^2,$$

因此 $\gamma_1,\cdots,\gamma_n$ 是 $\alpha_1,\cdots,\alpha_n$ 的一个排列，记之为 $\alpha_{\pi(1)},\cdots,\alpha_{\pi(n)}$，代入(3.6)，并利用引理 3.3，立即得到

$$\max_{U,V\in \mathscr{U}_n} \mathrm{Re}\ \mathrm{tr}(U\Lambda V\Omega) = \sum_{i=1}^{n} \alpha_i\beta_i. \tag{3.7}$$

因为等式(3.7)的两端是 $\alpha_1,\cdots,\alpha_n$ 与 $\beta_1,\cdots,\beta_n$ 的连续函数，而且该等式已在区域(3.4)内成立，所以取极限，立即推知等式(3.7)在区域

$$\alpha_1 \geqslant \cdots \geqslant \alpha_n \geqslant 0,\ \beta_1 \geqslant \cdots \geqslant \beta_n \geqslant 0$$

上亦成立. □

3.3 SG 函数

定义 3.2. 一个定义在 $\mathbb{R}^n$ 上的实值函数 Φ 叫做对称规度函数 (*the symmetric gauge function*, 以下简称为 SG 函数)，如果它满足下列条件:

I) $\boldsymbol{x} \neq 0 \Longrightarrow \Phi(\boldsymbol{x}) > 0$,

II) $\Phi(\rho\boldsymbol{x}) = |\rho|\Phi(\boldsymbol{x}),\ \forall\rho\in\mathbb{R}$,

III) $\Phi(\boldsymbol{x}+\boldsymbol{y}) \leqslant \Phi(\boldsymbol{x}) + \Phi(\boldsymbol{y})$,

IV) $\Phi(J\boldsymbol{x}_\pi) = \Phi(\boldsymbol{x})$,

V) $\Phi(\boldsymbol{e}_1) = 1$.

其中 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 是任意的实 n 维列向量，π 是 $1,\cdots,n$ 的任一排列，J 是对角线元素为 1 或 -1 的对角阵，$\boldsymbol{e}_1 = (1,0,\cdots,0)^T$.

对于 $\boldsymbol{x} = (\xi_1,\cdots,\xi_n)^T$，$\Phi(\boldsymbol{x})$ 记为

$$\Phi(\boldsymbol{x}) = \Phi(\xi_1,\cdots,\xi_n).$$

例 3.2. 容易验证，下述 $\Phi_k(\boldsymbol{x})(k=1,\cdots,n)$ 是 $\mathbb{R}^n$ 上的 SG 函数:

$$\Phi_k(\boldsymbol{x}) = \max_{1\leqslant i_1<\cdots<i_k\leqslant n}\{|\xi_{i_1}| + \cdots + |\xi_{i_k}|\},$$

$$\boldsymbol{x} = (\xi_1,\cdots,\xi_n)^T \in \mathbb{R}^n.$$

$\Phi_k(\boldsymbol{x})(k=1,\cdots,n)$叫做 $\mathbb{R}^n$ 上特殊的 SG 函数.

下面讨论 SG 函数的性质.

引理 3.4. 设 $\Phi(\boldsymbol{x})$ 是 $\mathbb{R}^n$ 上的 SG 函数, $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T$. 则对于 $0\leqslant p_i\leqslant 1$, 有

$$\Phi(p_1\xi_1,\cdots,p_n\xi_n)\leqslant\Phi(\xi_1,\cdots,\xi_n). \tag{3.8}$$

证明:

据定义 3.2 中的条件 IV), 无妨假定 $\xi_i\geqslant 0, i=1,\cdots,n$, 并且只需证明(3.8)式仅当 $p_1\neq 1$ 时成立即可, 即只需证明: 当 $0\leqslant p<1$ 时, 有

$$\Phi(p\xi_1,\xi_2,\cdots,\xi_n)\leqslant\Phi(\xi_1,\xi_2,\cdots,\xi_n).$$

利用定义 3.2,直接验证上述不等式如下:

$$\begin{aligned}\Phi(p\xi_1,\xi_2,\cdots,\xi_n)&=\Phi\left(\frac{1+p}{2}\xi_1+\frac{1-p}{2}(-\xi_1),\frac{1+p}{2}\xi_2\right.\\&\qquad\left.+\frac{1-p}{2}\xi_2,\cdots,\frac{1+p}{2}\xi_n+\frac{1-p}{2}\xi_n\right)\\&\leqslant\Phi\left(\frac{1+p}{2}(\xi_1,\xi_2,\cdots,\xi_n)\right)+\Phi\left(\frac{1-p}{2}(-\xi_1,\xi_2,\cdots,\xi_n)\right)\\&=\frac{1+p}{2}\Phi(\boldsymbol{x})+\frac{1-p}{2}\Phi(\boldsymbol{x})=\Phi(\boldsymbol{x}).\ \square\end{aligned}$$

引理 3.5. 设 $\Phi(\boldsymbol{x})$ 是 $\mathbb{R}^n$ 上的 SG 函数, $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T$. 则

$$\max_{1\leqslant i\leqslant n}|\xi_i|\leqslant\Phi(\boldsymbol{x})\leqslant\sum_{i=1}^n|\xi_i|.$$

证明:

据定义 3.2 中的条件 II),V) 及引理 3.4,有

$|\xi_i|=\Phi(0,\cdots,0,\xi_i,0,\cdots,0)\leqslant\Phi(\boldsymbol{x}), i=1,\cdots,n;$

另一方面,据定义 3.2 中的条件 II), III)和 V),有

$$\Phi(\boldsymbol{x})\leqslant\sum_{i=1}^n\Phi(0,\cdots,0,\xi_i,0,\cdots,0)\leqslant\sum_{i=1}^n|\xi_i|.\ \square$$

引理 3.6. SG 函数是 $\mathbb{R}^n$ 上的连续函数.

证明:

记 $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T$, $\boldsymbol{x}'=(\xi_1',\cdots,\xi_n')^T$. 根据定义 3.2 中

的条件 III)和引理 3.5，对于 SG 函数 Φ，必有

$$|\Phi(\boldsymbol{x}') - \Phi(\boldsymbol{x})| \leqslant \Phi(\boldsymbol{x}' - \boldsymbol{x}) \leqslant \sum_{i=1}^{n} |\xi_i' - \xi_i|.$$

由此可立即推知 $\Phi(\boldsymbol{x})$ 的连续性. □

由引理 3.6 可知，如果 $\Phi(\boldsymbol{x})$ 是 $\mathbb{R}^n$ 上的 SG 函数，则对于任一 $\boldsymbol{y} = (\eta_1, \cdots, \eta_n)^T \in \mathbb{R}^n$，$\sum_{i=1}^{n} \xi_i \eta_i \Big/ \Phi(\boldsymbol{x})$ 在有界闭集

$$\left\{\boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T \in \mathbb{R}^n: \sum_{i=1}^{n} |\xi_i| = 1\right\}$$

上连续，因而可以在其上达到极大值. 所以可以定义一个新的函数

$$\Psi(\boldsymbol{y}) = \max_{\sum_{i=1}^{n} |\xi_i| = 1} \frac{\sum_{i=1}^{n} \xi_i \eta_i}{\Phi(\boldsymbol{x})}, \quad \boldsymbol{y} = (\eta_1, \cdots, \eta_n)^T \in \mathbb{R}^n. \tag{3.9}$$

显然，$\Psi(\boldsymbol{y})$ 也可表示为

$$\Psi(\boldsymbol{y}) = \sup_{\boldsymbol{x} \neq 0} \frac{\sum_{i=1}^{n} \xi_i \eta_i}{\Phi(\boldsymbol{x})} = \sup_{\boldsymbol{x} \neq 0} \frac{\boldsymbol{x}^T \boldsymbol{y}}{\Phi(\boldsymbol{x})} \tag{3.10}$$

或

$$\Psi(\boldsymbol{y}) = \sup_{\Phi(\boldsymbol{x})=1} \sum_{i=1}^{n} \xi_i \eta_i = \sup_{\Phi(\boldsymbol{x})=1} \boldsymbol{x}^T \boldsymbol{y}. \tag{3.11}$$

可直接验证下述引理成立.

引理 3.7. 如果 $\Phi(\boldsymbol{x})$ 是 $\mathbb{R}^n$ 上的 SG 函数，则由(3.9)式定义的 $\Psi(\boldsymbol{y})$ 也是 $\mathbb{R}^n$ 上的 SG 函数.

因此引进下述定义:

定义 3.3. 设 $\Phi(\boldsymbol{x})$ 是 $\mathbb{R}^n$ 上的 SG 函数，则由(3.9)式定义的 $\Psi(\boldsymbol{y})$ 叫做 $\Phi(\boldsymbol{x})$ 的对偶 SG 函数.

由(3.10)立即得出

推论 3.1. 设 $\Phi(\boldsymbol{x})$ 是 $\mathbb{R}^n$ 上的 SG 函数，$\Psi(\boldsymbol{y})$ 是它的对偶

SG 函数，$x=(\xi_1,\cdots,\xi_n)^T$ 与 $y=(\eta_1,\cdots,\eta_n)^T\in \mathrm{R}^n$. 则

$$\sum_{i=1}^{n}\xi_i\eta_i\leqslant \Phi(x)\Psi(y). \tag{3.12}$$

引理 3.8. 设 Φ 是 R^n 上的 SG 函数，Ψ 是 Φ 的对偶 SG 函数. 则 Φ 亦是 Ψ 的对偶 SG 函数.

证明:

由(3.12)可知

$$\Phi(x)\geqslant \sup_{y\neq 0}\frac{x^Ty}{\Psi(y)}. \tag{3.13}$$

以下证明(3.13)中的等式成立.

令

$$\mathscr{B}=\{x\in \mathrm{R}^n:\Phi(x)\leqslant 1\}.$$

容易验证: $\mathscr{B}$ 是 R^n 中包含原点的一个凸体(所谓凸体,即指 R^n 中包含内点的有界闭凸集). 因而 (见[56] 第 8 章定理 3: "一个闭凸集等于包含它的所有半空间的交")

"如果一个点 x 属于 R^n 中每一个包含 $\mathscr{B}$ 的半空间

$$\{z\in \mathrm{R}^n:z^Ty\leqslant 1\},$$

则 $x\in\mathscr{B}$ (即 $\Phi(x)\leqslant 1$). "

由(3.10)和(3.11)可知

$$\mathscr{B}\subseteq\{z\in \mathrm{R}^n:z^Ty\leqslant 1\}\Longleftrightarrow \Psi(y)\leqslant 1. \tag{3.14}$$

事实上,(3.14)左端表示: $\Phi(x)\leqslant 1\Longrightarrow x^Ty\leqslant 1$. 特别地,对于满足 $\Phi(x)=1$ 的 x,必有 $x^Ty\leqslant 1$. 由(3.11)知 $\Psi(y)\leqslant 1$. 另一方面,据(3.10),由 $\Psi(y)\leqslant 1$ 可得 $x^Ty\leqslant\Phi(x)$. 因此,如果 $\Phi(x)\leqslant 1$, 则必有 $x^Ty\leqslant 1$. 即 $\mathscr{B}\subseteq\{z\in \mathrm{R}^n:z^Ty\leqslant 1\}$.

于是,上面引号" "中的断言可以改述为

"如果 x 满足'只要 $\Psi(y)\leqslant 1$ 就有 $x^Ty\leqslant 1$',

$$\text{则必有 }\Phi(x)\leqslant 1.\text{"} \tag{3.15}$$

注意到 $\Psi(y)$ 是 R^n 上的 SG 函数(引理 3.7),因而

$$\mathscr{B}'=\{y\in \mathrm{R}^n:\Psi(y)\leqslant 1\}$$

也是 R^n 中的一个包含原点的凸体. 于是对于任给的非零向量

$\boldsymbol{x} \in \mathbb{R}^n$，存在常数(仅与 $\boldsymbol{x}$ 有关)

$$p_{\boldsymbol{x}} = \max_{\boldsymbol{y} \in \mathscr{B}'} \boldsymbol{x}^T \boldsymbol{y} = \boldsymbol{x}^T \boldsymbol{y}_0, \boldsymbol{y}_0 \in \mathscr{B}'. \tag{3.16}$$

记 $\tilde{\boldsymbol{x}} = \dfrac{\boldsymbol{x}}{p_{\boldsymbol{x}}}$. (3.16)表明，只要 $\Psi(\boldsymbol{y}) \leqslant 1$，就有 $\tilde{\boldsymbol{x}}^T \boldsymbol{y} \leqslant 1$，并且存在 $\boldsymbol{y}_0 \in \mathscr{B}'$，使得 $\tilde{\boldsymbol{x}}^T \boldsymbol{y}_0 = 1$. 因此，根据(3.15)所述的结论，必有

$$\begin{aligned}\Phi(\tilde{\boldsymbol{x}}) \leqslant 1 = \max_{\boldsymbol{y} \in \mathscr{B}'} \tilde{\boldsymbol{x}}^T \boldsymbol{y} &\leqslant \sup_{\boldsymbol{y} \in \mathscr{B}' \setminus \{0\}} \frac{\tilde{\boldsymbol{x}}^T \boldsymbol{y}}{\Psi(\boldsymbol{y})} \\ &= \sup_{\boldsymbol{y} \neq 0} \frac{\tilde{\boldsymbol{x}}^T \boldsymbol{y}}{\Psi(\boldsymbol{y})}.\end{aligned}$$

再将 $\tilde{\boldsymbol{x}} = \dfrac{\boldsymbol{x}}{p_{\boldsymbol{x}}}$ 代入上式，即得

$$\Phi(\boldsymbol{x}) \leqslant \sup_{\boldsymbol{y} \neq 0} \frac{\boldsymbol{x}^T \boldsymbol{y}}{\Psi(\boldsymbol{y})}. \tag{3.17}$$

把(3.13)与(3.17)联系起来，便得到

$$\Phi(\boldsymbol{x}) = \sup_{\boldsymbol{y} \neq 0} \frac{\boldsymbol{x}^T \boldsymbol{y}}{\Psi(\boldsymbol{y})}.$$

即 Φ 是 Ψ 的对偶 SG 函数. □

定义 3.4. 设 $\Phi(\xi_1, \cdots, \xi_n)$ 是 $\mathbb{R}^n$ 上的 SG 函数. 对于任一矩阵 $A \in \mathbb{C}^{m \times n}$，如果 $\sigma(A) = \{\sigma_1, \cdots, \sigma_n\}$，则定义

$$\Phi(A) = \Phi(\sigma_1, \cdots, \sigma_n).$$

引理 3.9. 设 $A, X \in \mathbb{C}^{m \times n}$，$\Phi$ 与 Ψ 是 $\mathbb{R}^n$ 上互为对偶的 SG 函数. 则

$$\sup_{\Phi(X)=1} \operatorname{Re}\operatorname{tr}(X^H A) = \Psi(A).$$

证明:

设 $\sigma(X) = \{\xi_1, \cdots, \xi_n\}$, $\xi_1 \geqslant \cdots \geqslant \xi_n \geqslant 0$; $\sigma(A) = \{\sigma_1, \cdots, \sigma_n\}$, $\sigma_1 \geqslant \cdots \geqslant \sigma_n \geqslant 0$. 由

$$\Phi(X) = \Phi(U^H X V^H), \ \forall U \in \mathscr{U}_m, \ \forall V \in \mathscr{U}_n$$

可知，

$$\sup_{\Phi(X)=1} \operatorname{Re}\operatorname{tr}(X^H A) = \sup_{\Phi(U^H X V^H)=1} \operatorname{Re}\operatorname{tr}(V X^H U A)$$

$$= \sup_{\Phi(X)=1} \operatorname{Re} \operatorname{tr}(VX^HUA)$$

$$= \sup_{\Phi(X)=1} \max_{\substack{V \in \mathscr{U}_n \\ U \in \mathscr{U}_m}} \operatorname{Re} \operatorname{tr}(VX^HUA)$$

$$= \sup_{\Phi(\xi_1,\cdots,\xi_n)=1} \sum_{i=1}^{n} \xi_i \sigma_i \quad (\text{据定理 } 3.1)$$

$$= \Psi(\sigma_1, \cdots, \sigma_n) \ (\text{据}(3.11))$$

$$= \Psi(A) \ (\text{据定义 } 3.4). \ \square$$

引理 3.10. 设 $\boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T$ 与 $\boldsymbol{y} = (\eta_1, \cdots, \eta_n)^T$ 满足

$$\xi_1 \geqslant \cdots \geqslant \xi_n \geqslant 0, \quad \eta_1 \geqslant \cdots \geqslant \eta_n \geqslant 0. \tag{3.18}$$

则

$$\xi_1 + \cdots + \xi_k \leqslant \eta_1 + \cdots + \eta_k, \quad k = 1, \cdots, n \tag{3.19}$$

是使关系式

$$\Phi(\boldsymbol{x}) \leqslant \Phi(\boldsymbol{y}) \tag{3.20}$$

对 $\mathbf{R}^n$ 上每一个 SG 函数 Φ 均成立的必要与充分条件.

证明:

条件(3.19)的必要性.

利用 $\mathbf{R}^n$ 上特殊的 SG 函数

$$\Phi_k(\boldsymbol{x}) = \max_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \{|\xi_{i_1}| + \cdots + |\xi_{i_k}|\},$$

$$\boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T \in \mathbf{R}^n, \ k = 1, \cdots, n,$$

由

$$\Phi_k(\boldsymbol{x}) \leqslant \Phi_k(\boldsymbol{y}), \quad k = 1, \cdots, n$$

可立即导出分量为(3.18)的 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 必满足(3.19).

条件(3.19)的充分性.

1) 首先证明 A. S. Markus 的一个结果: 如果 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 的分量满足(3.18)与(3.19),则向量 $\boldsymbol{x}$ 必可表示为

$$\boldsymbol{x} = \sum_{l=1}^{N} \tau_l \boldsymbol{y}^{(l)} \quad (N = 2^n n!), \tag{3.21}$$

其中每个 $\boldsymbol{y}^{(l)}(l = 1, \cdots, N)$皆形如 $J\boldsymbol{y}_\pi$, $\boldsymbol{y}_\pi$ 表示 $\boldsymbol{y}$ 的分量经某一置换 π 所得到的向量, J 是对角线元素由 1 和 -1 组成的对角阵,

每个 $\tau_l \geqslant 0$ 并且 $\sum_{l=1}^{N} \tau_l = 1$.

(3.21) 的几何意义是：满足 (3.18) 与 (3.19) 的 $\boldsymbol{x}$，必属于 $\{\boldsymbol{y}^{(l)}\}_{l=1}^{N}$ 的凸包 K.

(3.21)可反证如下. 假定 $\boldsymbol{x} \notin K$，则必存在一个超平面

$$L = \left\{(\zeta_1, \cdots, \zeta_n)^T \in \mathbb{R}^n:\ \sum_{i=1}^{n} \alpha_i \zeta_i = \beta\right\} \subset \mathbb{R}^n,$$

使得K与 $\boldsymbol{x}$ 分别位于L的两侧. 更确切地说(如图 2-1)，有

$$\sum_{i=1}^{n} \alpha_i \zeta_i \leqslant \beta,\ \forall (\zeta_1, \cdots, \zeta_n)^T \in K$$

和

$$\sum_{i=1}^{n} \alpha_i \xi_i > \beta, \boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T.$$

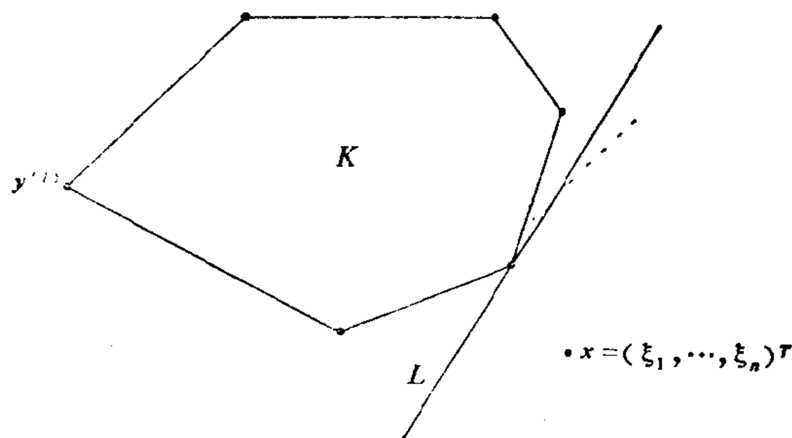


图 2-1

在 $\{\boldsymbol{y}^{(l)}\}_{l=1}^{N}$ 中，显然存在一个向量 $\boldsymbol{y}^{(p)} = (\eta_1^{(p)}, \cdots, \eta_n^{(p)})^T$，使得

$$\sum_{i=1}^{N} \alpha_i \eta_i^{(p)} = \sum_{i=1}^{n} \alpha'_i \eta_i,$$

其中 $\alpha'_1, \cdots, \alpha'_n$ 是 $|\alpha_1|, \cdots, |\alpha_n|$ 的一个排列，满足 $\alpha'_1 \geqslant \cdots \geqslant \alpha'_n$，因此，由 $y^{(p)} \in K$ 可知

$$\sum_{i=1}^{n} \alpha_i' \eta_i \leqslant \beta.$$

利用等式

$$\sum_{i=1}^{n} \alpha_i' \eta_i = \sum_{i=1}^{n} \alpha_n' \eta_i + \sum_{i=1}^{n-1} (\alpha_i' - \alpha_{i+1}') \sum_{j=1}^{i} \eta_j$$

和

$$\sum_{i=1}^{n} \alpha_i' \xi_i = \sum_{i=1}^{n} \alpha_n' \xi_i + \sum_{i=1}^{n-1} (\alpha_i' - \alpha_{i+1}') \sum_{j=1}^{i} \xi_j,$$

从(3.19)可得

$$\sum_{i=1}^{n} \alpha_i' \xi_i \leqslant \sum_{i=1}^{n} \alpha_i' \eta_i.$$

因此有

$$\sum_{i=1}^{n} \alpha_i' \xi_i \leqslant \beta. \tag{3.22}$$

但另一方面，由等式

$$\sum_{i=1}^{n} \alpha_i' \xi_i = \sum_{i=1}^{n} \alpha_i' \xi_n + \sum_{i=1}^{n} (\xi_i - \xi_{i+1}) \sum_{j=1}^{i} \alpha_j'$$

和

$$\sum_{i=1}^{n} \alpha_i \xi_i = \sum_{i=1}^{n} \alpha_i \xi_n + \sum_{i=1}^{n} (\xi_i - \xi_{i+1}) \sum_{j=1}^{i} \alpha_j$$

以及不等式

$$\sum_{i=1}^{k} \alpha_i' \geqslant \sum_{i=1}^{k} \alpha_i, \quad k = 1, \cdots, n$$

得到

$$\sum_{i=1}^{n} \alpha_i' \xi_i \geqslant \sum_{i=1}^{n} \alpha_i \xi_i > \beta. \tag{3.23}$$

(3.22)与(3.23)矛盾，所以(3.21)式成立.

2) 现在证明条件(3.19)的充分性. 设 Φ 是 $\mathbf{R}^n$ 上的 SG 函数. 利用表达式(3.21)和 SG 函数的性质 II、III、IV)，立即得出

$$\Phi(x) = \Phi\left(\sum_{l=1}^{N} \tau_l y^{(l)}\right) \leqslant \sum_{l=1}^{N} \tau_l \Phi(y^{(l)})$$

$$= \sum_{l=1}^{N} \tau_l \Phi(y) = \Phi(y). \quad \square$$

3.4 酉不变范数的性质

本节主要结果是下述定理.

定理 3.2 (von Neumann[128]). 设 Φ 是 $\mathbf{R}^n$ 上的 SG 函数，则 $\Phi(A)$ 是 $\mathbf{C}^{m\times n}$ 上的酉不变范数. 反之,对于 $\mathbf{C}^{m\times n}$ 上的每一个酉不变范数$\|\ \ \|$,必存在 $\mathbf{R}^n$ 上的 SG 函数 Φ,使得 $\|A\| = \Phi(A), \forall A \in \mathbf{C}^{m\times n}$.

证明:

设 Φ 是 $\mathbf{R}^n$ 上的 SG 函数, 现在证明 $\Phi(A)$ 是 $\mathbf{C}^{m\times n}$ 上的酉不变范数.

I) 显然 $\Phi(A) \geqslant 0$; 并且如果 $\Phi(A) = 0$, $\sigma(A) = \{\sigma_i\}$, 则有 $\Phi(\sigma_1, \cdots, \sigma_n) = 0$, 由此得到 $\sigma_1 = \cdots = \sigma_n = 0$, 从而 $A = 0$.

II) 因为 $\sigma(cA) = \{|c|\sigma_i\}$, $\forall c \in \mathbf{C}$, 所以

$$\Phi(cA) = \Phi(|c|\sigma_1, \cdots, |c|\sigma_n)$$
$$= |c|\Phi(\sigma_1, \cdots, \sigma_n) = |c|\Phi(A).$$

III) 据引理 3.9,有

$$\Phi(A+B) = \sup_{\Psi(X)=1} \mathrm{Re}\ \mathrm{tr}(X^H(A+B))$$
$$\leqslant \sup_{\Psi(X)=1} \mathrm{Re}\ \mathrm{tr}(X^H A) + \sup_{\Psi(X)=1} \mathrm{Re}\ \mathrm{tr}(X^H B)$$
$$= \Phi(A) + \Phi(B).$$

IV) 据定义 3.4,

$$\Phi(UAV) = \Phi(\sigma_1, \cdots, \sigma_n) = \Phi(A),$$
$$\forall U \in \mathscr{U}_m, \ \forall V \in \mathscr{U}_n.$$

V) 设 A 满足 $\mathrm{rank} A = 1$, 显然有 $\sigma(A) = \{\|A\|_2, 0, \cdots, 0\}$, 于是 $\Phi(A) = \Phi(\|A\|_2, 0, \cdots, 0) = \|A\|_2 \Phi(1, 0, \cdots, 0) = \|A\|_2$.

因此，$\Phi(A)$ 是 $\mathbb{C}^{m\times n}$ 上的酉不变范数.

反之，设$\| \ \|$是 $\mathbb{C}^{m\times n}$ 上的酉不变范数. 当$m\geqslant n$时，定义

$$\Phi(\xi_1,\cdots,\xi_n)=\left\|\begin{pmatrix}\xi_1 & & \\ & \ddots & \\ & & \xi_n \\ & 0 & \end{pmatrix}\right\|,\ (\xi_1,\cdots,\xi_n)^T\in\mathbb{R}^n.$$

直接验证可知，Φ 是 $\mathbb{R}^n$ 上的 SG 函数，而且显然有

$$\|A\|=\Phi(A),\ \forall A\in\mathbb{C}^{m\times n}.$$

当$m<n$时，首先在 $\mathbb{C}^{n\times m}$ 上定义非负实值函数$\| \ \|^*$:

$$\|B\|^*=\|B^T\|,\ \forall B\in\mathbb{C}^{n\times m}. \tag{3.24}$$

$\| \ \|^*$显然是 $\mathbb{C}^{n\times m}(n>m)$ 上的酉不变范数. 因此必存在 $\mathbb{R}^m$ 上的 SG 函数 Φ^*，使得 $\|B\|^*=\Phi^*(B)$, $\forall B\in\mathbb{C}^{n\times m}$.

在 $\mathbb{R}^n$ 上定义函数 Φ:

$$\Phi(x)=\Phi^*(\xi_{i_1},\cdots,\xi_{i_m}),x=(\xi_1,\cdots,\xi_n)^T\in\mathbb{R}^n,$$
$$|\xi_{i_1}|\geqslant|\xi_{i_2}|\geqslant\cdots\geqslant|\xi_{i_n}|.$$

容易证明：Φ 是 $\mathbb{R}^n$ 上的 SG 函数(本节习题 5).

注意到，当$m<n$时，如果 $A\in\mathbb{C}^{m\times n}$，$\sigma(A)=\{\sigma_i\}$ 满足 $\sigma_1\geqslant\cdots\geqslant\sigma_n\geqslant 0$，则必有

$$\sigma_1\geqslant\cdots\geqslant\sigma_m\geqslant 0=\sigma_{m+1}=\cdots=\sigma_n$$

和 $\sigma(A^T)=\{\sigma_1,\cdots,\sigma_m\}$. 因此，有

$$\begin{aligned}\|A\|&=\|A^T\|^*=\Phi^*(A^T)=\Phi^*(\sigma_1,\cdots,\sigma_m)\\&=\Phi(\sigma_1,\cdots,\sigma_m,0,\cdots,0)\\&=\Phi(\sigma_1,\cdots,\sigma_m,\sigma_{m+1},\cdots,\sigma_n)\\&=\Phi(A).\ \square\end{aligned}$$

此外，酉不变范数还有下面一些重要性质.

定理 3.3. 设 $A,B\in\mathbb{C}^{m\times n}$，它们的奇异值分别为 $\sigma_1\geqslant\cdots\geqslant\sigma_n\geqslant 0$ 和 $\tau_1\geqslant\cdots\geqslant\tau_n\geqslant 0$. 如果 $\sigma_i\leqslant\tau_i, i=1,\cdots,n$，则在 $\mathbb{C}^{m\times n}$ 上的任一酉不变范数$\| \ \|$，必有 $\|A\|\leqslant\|B\|$.

证明:

据定理 3.2，存在 $\mathbb{R}^n$ 上的 SG 函数 Φ，使得 $\|A\|=\Phi(A)$, $\forall A\in\mathbb{C}^{m\times n}$. 由定义 3.4 知，只需证明 $\Phi(\sigma_1,\cdots,\sigma_n)\leqslant\Phi(\tau_1,\cdots,$

τ_n). 而这一不等式可由 $\sigma_i \leqslant \tau_i(1 \leqslant i \leqslant n)$ 及引理 3.4 立即得到. □

定理 3.4. 设$\|\ \ \|$为 $\mathbf{C}^{m\times n}$ 上的酉不变范数,则有

$$\|AB\| \leqslant \|A\|_2\|B\|, \forall A \in \mathbf{C}^{m\times m}, \forall B \in \mathbf{C}^{m\times n} \tag{3.25}$$

和

$$\|AB\| \leqslant \|A\|\|B\|_2, \forall A \in \mathbf{C}^{m\times n}, \forall B \in \mathbf{C}^{n\times n}. \tag{3.26}$$

证明:

设 $A \in \mathbf{C}^{m\times m}, B \in \mathbf{C}^{m\times n}$. 由

$$B^H A^H A B \leqslant \|A\|_2^2 B^H B$$

和 Hermite 阵的特征值的 Minimax 性质(见第三章 § 3),可以推知,如果

$$\sigma(AB) = \{\gamma_1, \cdots, \gamma_n\}, \sigma(B) = \{\tau_1, \cdots, \tau_n\},$$

其中 $\gamma_1 \geqslant \cdots \geqslant \gamma_n \geqslant 0$, $\tau_1 \geqslant \cdots \geqslant \tau_n \geqslant 0$, 则

$$\gamma_i \leqslant \|A\|_2 \tau_i, i = 1, \cdots, n. \tag{3.27}$$

设 $\mathbf{R}^n$ 上与$\|\ \ \|$相对应的 SG 函数为 Φ. 利用定理 3.3 和 (3.27) 得到

$$\begin{aligned}\|AB\| &= \Phi(AB) = \Phi(\gamma_1, \cdots, \gamma_n) \\ &\leqslant \Phi(\|A\|_2\tau_1, \cdots, \|A\|_2\tau_n) \\ &= \|A\|_2\Phi(\tau_1, \cdots, \tau_n) = \|A\|_2\|B\|.\end{aligned}$$

即不等式(3.25)成立.

同理可证不等式 (3.26). □

定理 3.5. $\mathbf{C}^{m\times n}$ 上的酉不变范数必与向量的 Euclid 范数相容,并且

$$\|A\|_2 \leqslant \|A\| \leqslant r\|A\|_2, \forall A \in \mathbf{C}^{m\times n},$$

其中 $r = \min\{m, n\}$.

证明:

任取一非零向量 $\boldsymbol{x} \in \mathbf{C}^n$,构造矩阵$(\boldsymbol{x}, 0, \cdots, 0) \in \mathbf{C}^{n\times n}$. 于是根据定理 3.4,对于任一矩阵 $A \in \mathbf{C}^{m\times n}$, 有

$$\begin{aligned}\|A\boldsymbol{x}\|_2 &= \|A(\boldsymbol{x}, 0, \cdots, 0)\| \leqslant \|A\|\|(\boldsymbol{x}, 0, \cdots, 0)\|_2 \\ &= \|A\|\|\boldsymbol{x}\|_2,\end{aligned}$$

即$\| \quad \|$与向量的 Euclid 范数相容．应用定理 2.8，有

$$\|A\|_2 \leqslant \|A\|.$$

另一方面，据定理 3.2，存在 R^n 上的 SG 函数 Φ，使得

$$\|A\| = \Phi(\sigma_1, \cdots, \sigma_n),$$

其中 $\sigma_1, \cdots, \sigma_n$ 是 A 的奇异值．于是有

$$\|A\| \leqslant \sum_{i=1}^{n} \sigma_i \Phi(1, 0, \cdots, 0) = \sum_{i=1}^{n} \sigma_i$$
$$\leqslant \min\{m, n\} \cdot \max_i \sigma_i = r\|A\|_2. \quad \square$$

由定理 3.4 与定理 3.5 立即得到

推论 3.2. $C^{n\times n}$ 上的酉不变范数是相容范数．即：如果$\| \quad \|$是 $C^{n\times n}$ 上的酉不变范数，则

$$\|AB\| \leqslant \|A\|\|B\|, \ \forall A, \ B \in C^{n\times n}.$$

例 3.3. 由例 3.2 和定理 3.2 可知，对于任一 $A \in C^{m\times n}$，如果 $\sigma(A) = \{\sigma_i\}$ 满足 $\sigma_1 \geqslant \cdots \geqslant \sigma_n \geqslant 0$，则由

$$\|A\|_{(k)} = \sigma_1 + \cdots + \sigma_k, \ k = 1, \cdots, n, \ \forall A \in C^{m\times n}$$

所定义的$\| \|_{(k)}$是 $C^{m\times n}$ 上的酉不变范数．通常把$\| \|_{(k)}$叫做 $C^{m\times n}$ 上的特殊的酉不变范数．显然有

$$\|A\|_{(1)} = \|A\|_2, \ \forall A \in C^{m\times n}.$$

下述定理说明了特殊的酉不变范数的重要意义．它是引理 3.10 的直接推论．

定理 3.6. 设A，$B \in C^{m\times n}$．则

$$\|A\| \leqslant \|B\| \tag{3.28}$$

对于$C^{m\times n}$上所有的酉不变范数均成立的必要与充分条件是(3.28)式对于每个特殊的酉不变范数$\| \quad \|_{(k)}$均成立，即

$$\|A\|_{(k)} \leqslant \|B\|_{(k)}, \ k = 1, \cdots, n.$$

习题

1．证明定理 3.1 当 $m > n$ 时亦成立．

2．设 $A, B \in C^{m\times n}$，$\sigma(A) = \{\alpha_i\}$与$\sigma(B) = \{\beta_i\}$满足 $\alpha_1 \geqslant \cdots \geqslant \alpha_n \geqslant 0$ 与 $\beta_1 \geqslant \cdots \geqslant \beta_n \geqslant 0$．试利用定理 3.1 证明

$$\|A-B\|_F^2 \geqslant \sum_{i=1}^{n}(\alpha_i-\beta_i)^2.$$

3. 设 $A=U\Sigma V^H$ 是 $A\in\mathbf{C}^{n\times n}$ 的奇异值分解，其中 $U,V\in\mathscr{U}_n$，$\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_n)>0$. 试证

$$\mathrm{Re}\ \mathrm{tr}(U\Sigma V^H)\leqslant \mathrm{tr}(\Sigma),$$

并且等式成立的必要与充分条件是 $U=V$. 进而证明：当且仅当 $AW>0$ 时，$\max\limits_{W\in\mathscr{U}_n}\mathrm{Re}\ \mathrm{tr}(AW)$ 达到.

4. 设 $A,H\in\mathbf{C}^{n\times n}$. 若 H 为 Hermite 阵，则对于 $\mathbf{C}^{n\times n}$ 上的任一酉不变范数 $\|\ \ \|$，有

$$\left\|A-\frac{A+A^H}{2}\right\|\leqslant\|A-H\|.$$

5. 设 $\Phi^*(\boldsymbol{y})$ 是 $\mathbf{R}^m$ 上的 SG 函数. 在 $\mathbf{R}^n(n>m)$ 上定义函数：

$$\Phi(\boldsymbol{x})=\Phi^*(\xi_{i_1},\cdots,\xi_{i_m}),\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T\in\mathbf{R}^n,$$
$$|\xi_{i_1}|\geqslant\cdots\geqslant|\xi_{i_n}|.$$

试证 Φ 是 $\mathbf{R}^n$ 上的 SG 函数.

6. 在 $\mathbf{C}^{n\times n}$ 上定义非负实值函数 ν_k：对任一 $A\in\mathbf{C}^{n\times n}$，如果 $\sigma(A)=\{\sigma_i(A)\}$ 满足 $\sigma_1(A)\geqslant\cdots\geqslant\sigma_n(A)\geqslant 0$，则令

$$\nu_k(A)=[\sigma_1^2(A)+\cdots+\sigma_k^2(A)]^{\frac{1}{2}},\ k=1,\cdots,n.$$

试证 $\nu_k(k=1,2,\cdots,n)$ 是 $\mathbf{C}^{n\times n}$ 上的酉不变范数.

7. 设 $C=A+B\in\mathbf{C}^{m\times n}$，$\sigma(A)=\{\alpha_i\}$，$\sigma(B)=\{\beta_i\}$ 与 $\sigma(C)=\{\gamma_i\}$ 满足

$$\alpha_1\geqslant\cdots\geqslant\alpha_n\geqslant 0,$$
$$\beta_1\geqslant\cdots\geqslant\beta_n\geqslant 0$$

和

$$\gamma_1\geqslant\cdots\geqslant\gamma_n\geqslant 0.$$

试证

$$\sum_{i=1}^{k}\gamma_i\leqslant\sum_{i=1}^{k}\alpha_i+\sum_{i=1}^{k}\beta_i,\ k=1,\cdots,n.$$

8. 设 $A\in\mathbf{C}^{n\times n}$. 试证

$$\frac{1}{2}\|A\|_2 \leqslant \sup_{\|u\|_2=1} |u^H A u| \leqslant \|A\|_2.$$

§4 $\mathbb{C}_l^n$ 上的度量

4.1 基本概念

$\mathbf{C}^n$ 内所有 l 维列空间的全体，记作 $\mathbf{G}_l^n$.

$\mathbf{G}_l^n$ 也可以如下定义. 首先把 $\mathbf{C}_l^{n\times l}$ 内的点划分成等价类：如果存在非奇异阵 $Q\in \mathbf{C}^{l\times l}$，使得 $Z_1 = Z_2Q$，就称点 Z_1 与 Z_2 属于同一个等价类(记作 $Z_1 \sim Z_2$)；把 $\mathbf{C}_l^{n\times l}$ 内的每一个等价类看作一点，由所有这种点构成的空间，就是 $\mathbf{G}_l^n$. 通常把 $\mathbf{G}_l^n$ 叫做复投影空间(或 Grassmann 流形)，它的每一个点可以用 $R(Z)$ 表示，其中 $Z\in \mathbf{C}_l^{n\times l}$；也可以简单地用 Z 表示 $\mathbf{G}_l^n$ 中的一个点，这时 Z 是由一个 l 维子空间中任取一组基底作为列向量构成的矩阵，或者说，Z 是一个等价类中的代表元素.

$\mathbf{C}^n$ 内所有 l 维行空间的全体，记作 $\mathbf{G}_{l,n}$. 仿照上面的办法，可以通过在 $\mathbf{C}_l^{l\times n}$ 内划分等价类的办法，定义复投影空间 $\mathbf{G}_{l,n}$.

以下仅讨论 $\mathbf{G}_l^n$. 所得到的结论，同样适用于 $\mathbf{G}_{l,n}$.

定义 4.1. 一个定义在 $\mathbf{G}_l^n$ 上的实值函数 $d(\cdot,\ \cdot)$ 叫做 $\mathbf{G}_l^n$ 上的度量(或者叫距离)，如果它满足

I) $d(\mathscr{Z},\mathscr{W}) \geqslant 0$，并且 $d(\mathscr{Z},\mathscr{W}) = 0 \Longleftrightarrow \mathscr{Z} = \mathscr{W}$；

II) $d(\mathscr{Z},\mathscr{W}) = d(\mathscr{W},\mathscr{Z})$；

III) $d(\mathscr{Z},\mathscr{W}) \leqslant d(\mathscr{Z},\mathscr{X}) + d(\mathscr{X},\mathscr{W})$.

其中 $\mathscr{Z}$，$\mathscr{W}$ 与 $\mathscr{X}$ 是 $\mathbf{G}_l^n$ 内任意的点.

定义 4.2. 设 $d(\cdot,\cdot)$是一个定义在 $\mathbf{G}_l^n$ 上的实值函数. 如果它除了满足定义 4.1 中的三条性质 I)、II)、III) 之外，还满足

IV) $d(U\mathscr{Z},U\mathscr{W}) = d(\mathscr{Z},\mathscr{W}), \forall U\in \mathscr{U}_n$，

则称 $d(\cdot,\cdot)$ 是 $\mathbf{G}_l^n$ 上的酉不变度量.

如何在 $\mathbf{G}_l^n$ 上引进度量，是一个比较复杂的问题. 首先考察一个最简单的情形.

今在 $\mathbb{R}^2$ 内任取二向量 $\boldsymbol{x}=(\xi_1,\xi_2)^T$ 与 $\boldsymbol{y}=(\eta_1,\eta_2)^T$. 假设它们之间的夹角是 $\varphi(0\leqslant\varphi\leqslant\pi)$，则有关系式

$$\cos\varphi=\frac{\boldsymbol{y}^T\boldsymbol{x}}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2},$$

其中 $\|\ \|_2$ 表示向量的 Euclid 范数.但是如果考虑两个子空间 $R(\boldsymbol{x})$ 与 $R(\boldsymbol{y})$，假设它们之间的夹角为 $\theta\left(0\leqslant\theta\leqslant\frac{\pi}{2}\right)$,则应有

$$\cos\theta=\frac{|\boldsymbol{y}^T\boldsymbol{x}|}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2},$$

即

$$\theta(R(\boldsymbol{x}),\ R(\boldsymbol{y}))=\arccos\frac{|\boldsymbol{y}^T\boldsymbol{x}|}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2}\tag{4.1}$$

(见图 2-2).

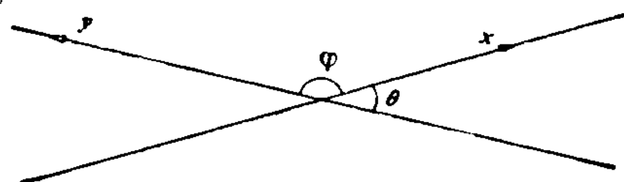


图 2-2

例如，$\varphi(\boldsymbol{x},-\boldsymbol{x})=\pi$，但 $\theta(R(\boldsymbol{x}),R(-\boldsymbol{x}))=0$. 因此，$\mathbb{R}^2$ 中两个一维子空间之间的夹角与两个向量之间的夹角，在概念上是不同的.

在 $\mathbb{R}^2$ 内任给二子空间 $R(\boldsymbol{x})$ 与 $R(\boldsymbol{y})$，显然可以用它们之间的夹角 $\theta(R(\boldsymbol{x}),\ R(\boldsymbol{y}))$ 作为它们之间的距离. 容易验证，$\sin\theta(R(\boldsymbol{x}),\ R(\boldsymbol{y}))$ 也满足定义 4.1 中的三条性质(本节习题 1)，因而也可以用来作为 $R(\boldsymbol{x})$ 与 $R(\boldsymbol{y})$ 之间的距离.

值得指出的是，有

$$\sin\theta(R(\boldsymbol{x}),R(\boldsymbol{y}))=\|P_{R(\boldsymbol{x})}-P_{R(\boldsymbol{y})}\|_2.\tag{4.2}$$

其中 $P_{R(\boldsymbol{x})}$ 与 $P_{R(\boldsymbol{y})}$ 分别表示到 $R(\boldsymbol{x})$ 与 $R(\boldsymbol{y})$ 上的正交投影算子(以下简称投影). (4.2) 式可以如下证明：利用旋转，无妨假设

$R(\boldsymbol{x}) = R\left(\begin{pmatrix}1\\0\end{pmatrix}\right)$，并在 $R(\boldsymbol{y})$ 内取一基底 $\begin{pmatrix}\cos\theta\\\sin\theta\end{pmatrix}$，见图 2-3.
于是有

$$\|P_{R(\boldsymbol{x})} - P_{R(\boldsymbol{y})}\|_2 = \left\|\begin{pmatrix}1\\0\end{pmatrix}(1,0) - \begin{pmatrix}\cos\theta\\\sin\theta\end{pmatrix}(\cos\theta, \sin\theta)\right\|_2$$

$$= \left\|\begin{pmatrix}\sin^2\theta & -\sin\theta\cos\theta\\ -\sin\theta\cos\theta & -\sin^2\theta\end{pmatrix}\right\|_2 = \sin\theta.$$

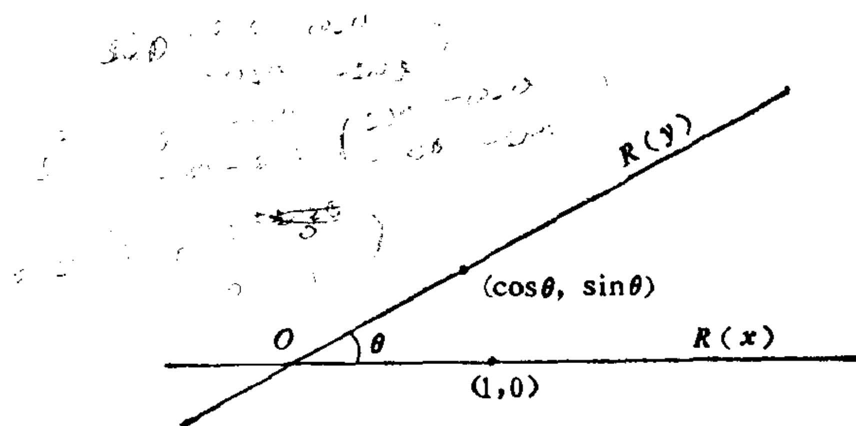


图 2-3

(4.2) 表明，$\|P_{R(\boldsymbol{x})} - P_{R(\boldsymbol{y})}\|_2$ 是 $\mathbb{R}^2$ 内任意两个一维子空间 $R(\boldsymbol{x})$ 与 $R(\boldsymbol{y})$ 之间的距离. 由此启发了一个一般性的命题:

定理 4.1. 如果$\|\ \ \|$是 $\mathbf{C}^{n\times n}$ 上的范数，则

$$\|P_{R(Z)} - P_{R(W)}\|$$

是 $\mathbf{G}_l^n$ 上的度量，其中 $Z, W \in \mathbf{C}_l^{n\times l}$；如果$\|\ \ \|$是 $\mathbf{C}^{n\times n}$ 上的酉不变范数，则 $\|P_{R(Z)} - P_{R(W)}\|$ 是 $\mathbf{G}_l^n$ 上的酉不变度量.

定理 4.1 的证明，作为一个练习(本节习题 2).

4.2 关于 $\|\sin\Theta(Z,W)\|_2$

(4.1)所示的夹角，可以推广到一般的情况.

设 $Z, W \in \mathbf{C}_l^{n\times l}$，令

$$Z_1 = Z(Z^HZ)^{-\frac{1}{2}},\ W_1 = W(W^HW)^{-\frac{1}{2}}. \tag{4.3}$$

定义

$$\Theta(Z, W) = \arccos(Z_1^HW_1W_1^HZ_1)^{\frac{1}{2}} \geqslant 0. \tag{4.4}$$

(4.4) 式所示的 $\Theta(Z,W)\in \mathbf{C}^{l\times l}$，文献中往往记作 $\Theta(R(Z),R(W))$.

下述定理是(4.2)的推广.

定理 4.2. 对于谱范数，有等式

$$\|\sin\Theta(Z,W)\|_2=\|P_Z-P_W\|_2, \forall Z,W\in \mathbf{C}_l^{n\times l}. \tag{4.5}$$

证明:

取 Z_1 与 W_1 如(4.3)所示. 显然有

$$\sin\Theta(Z,W)=\sin\Theta(Z_1,W_1)$$

和

$$P_Z=P_{Z_1},\ P_W=P_{W_1}.$$

因此无妨假设 Z 与 W 满足 $Z^HZ=W^HW=I$.

根据第一章定理 3.5，存在 $Q\in\mathscr{U}_n$ 与 $U_1,V_1\in\mathscr{U}_l$，使得当 $2l\leqslant n$ 时，有

$$QZU_1=\begin{pmatrix} I\\ 0\\ 0\end{pmatrix}\begin{matrix} l\\ l\\ n-2l\end{matrix},\quad QWV_1=\begin{pmatrix} \Gamma\\ \Sigma\\ 0\end{pmatrix}\begin{matrix} l\\ l\\ n-2l\end{matrix},$$
$$\qquad\qquad l \qquad\qquad\qquad\qquad l$$

其中 $\Gamma=\mathrm{diag}(\gamma_1,\cdots,\gamma_l)$，$\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_l)$，$0\leqslant\gamma_1\leqslant\cdots\leqslant\gamma_l$，$\sigma_1\geqslant\cdots\geqslant\sigma_l\geqslant 0,\gamma_i^2+\sigma_i^2=1,i=1,\cdots,l$; 当 $2l\geqslant n$ 时，有

$$QZU_1=\begin{pmatrix} I & 0\\ 0 & I\\ 0 & 0\end{pmatrix}\begin{matrix} n-l\\ 2l-n\\ n-l\end{matrix},$$
$$n-l\ \ 2l-n$$

$$QWV_1=\begin{pmatrix} \Gamma & 0\\ 0 & I\\ \Sigma & 0\end{pmatrix}\begin{matrix} n-l\\ 2l-n\\ n-l\end{matrix},$$
$$n-l\ \ 2l-n$$

其中 $\Gamma=\mathrm{diag}(\gamma_1,\cdots,\gamma_{n-l})$，$\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_{n-l})$，$0\leqslant\gamma_1\leqslant\cdots\leqslant\gamma_{n-l}$，$\sigma_1\geqslant\cdots\geqslant\sigma_{n-l}\geqslant 0$，$\gamma_i^2+\sigma_i^2=1$，$i=1,\cdots,$

$n-l$.

于是，当 $2l \leqslant n$ 时，得到

$$\|\sin\Theta(Z, W)\|_2 = \|I - Z^H W W^H Z\|_2^{\frac{1}{2}}$$
$$= \|I - U_1^H \Gamma^2 U_1\|_2^{\frac{1}{2}} = \|U_1^H (I - \Gamma^2) U_1\|_2^{\frac{1}{2}}$$
$$= \|I - \Gamma^2\|_2^{\frac{1}{2}} = \sigma_1;$$

当 $2l \geqslant n$ 时，得到

$$\|\sin\Theta(Z,W)\|_2 = \left\|I - U_1^H \begin{pmatrix} \Gamma^2 & 0 \\ 0 & I \end{pmatrix} U_1\right\|_2^{\frac{1}{2}}$$
$$= \left\|U_1^H \begin{pmatrix} I - \Gamma^2 & 0 \\ 0 & 0 \end{pmatrix} U_1\right\|_2^{\frac{1}{2}} = \left\|\begin{pmatrix} I - \Gamma^2 & 0 \\ 0 & 0 \end{pmatrix}\right\|_2^{\frac{1}{2}}$$
$$= \sigma_1.$$

另一方面，当 $2l \leqslant n$ 时，有

$$\|P_Z - P_W\|_2 = \|ZZ^H - WW^H\|_2$$
$$= \left\|Q^H \begin{pmatrix} I - \Gamma^2 & -\Gamma\Sigma & 0 \\ -\Sigma\Gamma & -\Sigma^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} Q\right\|_2$$
$$= \left\|\begin{pmatrix} \Sigma^2 & -\Gamma\Sigma & 0 \\ -\Sigma\Gamma & -\Sigma^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}\right\|_2$$
$$= \sigma_1; \tag{4.6}$$

当 $2l \geqslant n$ 时，有

$$\|P_Z - P_W\|_2 = \left\|\begin{pmatrix} \Sigma^2 & 0 & -\Gamma\Sigma \\ 0 & 0 & 0 \\ -\Sigma\Gamma & 0 & -\Sigma^2 \end{pmatrix}\right\|_2 = \sigma_1.$$

比较上列诸式，便得到(4.5). □

另外一种描述两个子间 $\mathscr{Z} = R(Z)$ 与 $\mathscr{W} = R(W)$ 接近程度的办法，是考虑它们之间的裂口(*gap*)，即

$$\gamma(\mathscr{Z},\mathscr{W})$$

$$=\max\left\{\sup_{\substack{\|z\|=1\\ z\in\mathscr{Z}}}\inf_{w\in\mathscr{W}}\|w-z\|,\sup_{\substack{\|w\|=1\\ w\in\mathscr{W}}}\inf_{z\in\mathscr{Z}}\|z-w\|\right\}. \quad (4.7)$$

其中$\|\ \|$是任一种向量范数.

定理 4.3. 如果 (4.7) 式中用向量的 Euclid 范数 $\|\ \|_2$ 定义 $\gamma(\mathscr{Z},\mathscr{W})$，并记作 $\gamma_2(\mathscr{Z},\mathscr{W})$，则

$$\gamma_2(\mathscr{Z},\mathscr{W})=\|P_Z-P_W\|_2,\quad \forall Z,W\in\mathbb{C}_l^{n\times l}, \quad (4.8)$$

其中 $\mathscr{Z}=R(Z),\mathscr{W}=R(W)$.

证明:

以下证明当 $2l\leqslant n$ 的情形（当 $2l\geqslant n$ 时，同理可证）. 分两步进行.

1）根据第一章定理 3.5，利用适当的酉变换，并适当选取标准正交基底，无妨假定

$$\mathscr{Z}=R\left(\begin{pmatrix}I^{(l)}\\0\\0\end{pmatrix}\right),\quad \mathscr{W}=R\left(\begin{pmatrix}\Gamma\\\Sigma\\0\end{pmatrix}\right), \quad (4.9)$$

其中 Γ 与 Σ 如第一章(3.11)与(3.12)式所示. 令

$$U=\begin{pmatrix}\Gamma & \Sigma & 0\\ \Sigma & -\Gamma & 0\\ 0 & 0 & I\end{pmatrix}\in\mathbb{C}^{n\times n},$$

U 显然是一酉阵，并且满足

$$U\begin{pmatrix}I^{(l)}\\0\\0\end{pmatrix}=\begin{pmatrix}\Gamma\\\Sigma\\0\end{pmatrix},\quad U\begin{pmatrix}\Gamma\\\Sigma\\0\end{pmatrix}=\begin{pmatrix}I\\0\\0\end{pmatrix}. \quad (4.10)$$

(4.10)表示:

$$U\mathscr{Z}=\mathscr{W},\ U\mathscr{W}=\mathscr{Z}. \quad (4.11)$$

于是有

$$\sup_{\substack{\|z\|_2=1\\ z\in\mathscr{Z}}}\inf_{w\in\mathscr{W}}\|w-z\|_2=\sup_{\substack{\|w'\|_2=1\\ w'\in\mathscr{W}}}\inf_{z'\in\mathscr{Z}}\|Uz'-Uw'\|_2$$

$$=\sup_{\substack{\|w\|_2=1\\ w\in\mathscr{W}}}\inf_{z\in\mathscr{Z}}\|z-w\|_2,$$

因此，

$$\gamma_2(\mathscr{Z},\mathscr{W})=\sup_{\substack{\|z\|_2=1\\ z\in\mathscr{Z}}}\ \inf_{w\in\mathscr{W}}\|w-z\|. \tag{4.12}$$

2) 由(4.9)和(4.12)知

$$\begin{aligned}\gamma_2^2(\mathscr{Z},\mathscr{W})&=\sup_{\substack{\|u\|_2=1\\ u\in\mathbb{C}^l}}\ \inf_{v\in\mathbb{C}^l}\left\|\begin{pmatrix}\Gamma\\ \Sigma\\ 0\end{pmatrix}v-\begin{pmatrix}I\\ 0\\ 0\end{pmatrix}u\right\|_2^2\\ &=\sup_{\substack{\|u\|_2=1\\ u\in\mathbb{C}^l}}\ \inf_{v\in\mathbb{C}^l}\{\|u-\Gamma v\|_2^2+\|\Sigma v\|_2^2\}\\ &=\sup_{\substack{\|u\|_2=1\\ u\in\mathbb{C}^l}}\ \inf_{v\in\mathbb{C}^l}\{1+\|v\|_2^2-2\mathrm{Re}\,u^H\Gamma v\}.\end{aligned} \tag{4.13}$$

在(4.13)式右端特别地取 $v=\Gamma u$，并利用(4.6)，得到

$$\begin{aligned}\gamma_2^2(\mathscr{Z},\mathscr{W})&\leqslant\sup_{\substack{\|u\|_2=1\\ u\in\mathbb{C}^l}}(1-\|\Gamma u\|_2^2)=1-\gamma_1^2\\ &=\sigma_1^2=\|P_Z-P_W\|_2^2;\end{aligned} \tag{4.14}$$

另一方面，在 (4.13) 中特别地取 $u=(1,0,\cdots,0)^T$，并记 $v=(\nu_1,\cdots,\nu_l)^T$，得到

$$\begin{aligned}\gamma_2^2(\mathscr{Z},\mathscr{W})&\geqslant\inf_{v\in\mathbb{C}^l}(1+\|v\|_2^2-2\gamma_1\mathrm{Re}\,\nu_1)\\ &\geqslant\inf_{\nu_1\in\mathbb{C}}[1+(\mathrm{Re}\,\nu_1)^2-2\gamma_1\mathrm{Re}\,\nu_1]\\ &=\inf_{t\in\mathbb{R}}(t^2-2\gamma_1t+1)=1-\gamma_1^2\\ &=\sigma_1^2=\|P_Z-P_W\|_2^2.\end{aligned} \tag{4.15}$$

把(4.14)与(4.15)联系起来，便得到(4.8)式. □

从定理4.2和定理4.3的证明可以看出 $\|\sin\Theta(Z,W)\|_2$ 的几何意义：在经过适当坐标变换，并在变换之后的 $\mathscr{Z}$ 与 $\mathscr{W}$ 中适当选取标准正交基底之后，这两组基底的对应向量之间的夹角的余弦即为 $\gamma_i=\cos\theta_i$，$i=1,\cdots,l$，它们满足 $0\leqslant\gamma_1\leqslant\cdots\leqslant\gamma_l$，即 $\frac{\pi}{2}\geqslant\theta_1\geqslant\cdots\geqslant\theta_l\geqslant0$. 而 $\|\sin\Theta(Z,W)\|_2$ 恰好是它们之间最大夹角 θ_1 的正弦 $\sin\theta_1=\sigma_1$.

此外，如果在 $\mathbb{C}^n$ 中的两个 l 维列空间 $\mathscr{Z}$ 与 $\mathscr{W}$ 中分别选取标

准正交基底构成 Z 与 W，利用第一章定理 3.5，还可以直接验证(作为本节习题 4)：

$$\|\sin\Theta(Z,W)\|_2 = \|(I-P_W)Z\|_2 = \|(I-P_Z)W\|_2$$
$$= \|(I-P_W)P_Z\|_2 = \|(I-P_Z)P_W\|_2. \tag{4.16}$$

4.3 关于 $\|\sin\Theta(Z,W)\|$

定理 4.2 表明，$\|\sin\Theta(Z,W)\|_2$ 是 $\mathbf{G}_l^n$ 上的度量. 但是当$\|\ \ \|$是 $\mathbf{C}^{l\times l}$ 上的任一范数时，$\|\sin\Theta(Z,W)\|$还是不是度量？下述定理对于$\|\ \ \|$是 $\mathbf{C}^{l\times l}$ 上的酉不变范数的情形，给予了肯定的回答.

定理 4.4.[18] 设$\|\ \ \|$是 $\mathbf{C}^{l\times l}$ 上的酉不变范数，则 $\|\sin\Theta(Z,W)\|$必是 $\mathbf{G}_l^n$ 上的酉不变度量.

证明：

设 $2l\leqslant n$(当 $2l>n$ 时，同理可证). 据第一章定理 3.5，利用适当酉变换，并在 $R(Z)$与 $R(W)$ 中适当选取标准正交基底，可以假设

$$Z=\begin{pmatrix} I\\ 0\\ 0\end{pmatrix},\quad W=\begin{pmatrix}\Gamma\\ \Sigma\\ 0\end{pmatrix},$$

其中 $\Gamma=\mathrm{diag}(\gamma_1,\cdots,\gamma_l)\geqslant 0$, $\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_l)\geqslant 0$. 于是有

$$\sin\Theta(Z,W)=\mathrm{diag}(\sigma_1,\cdots,\sigma_l) \tag{4.17}$$

和

$$P_Z-P_W=U\mathrm{diag}(\sigma_1,\sigma_1,\cdots,\sigma_l,\sigma_l,0^{(n-2l)})V, \tag{4.18}$$

其中

$$U=P\mathrm{diag}\left(\begin{pmatrix}\sigma_1 & \gamma_1\\ -\gamma_1 & \sigma_1\end{pmatrix},\cdots,\begin{pmatrix}\sigma_l & \gamma_l\\ -\gamma_l & \sigma_l\end{pmatrix}, I^{(n-2l)}\right),$$
$$V=\mathrm{diag}(1,-1,\cdots,1,-1,I^{(n-2l)})P.$$

P 是一排列方阵，显然 $U,V\in\mathscr{U}_n$.

设$\|\ \ \|$是 $\mathbf{C}^{l\times l}$ 上的酉不变范数，则据定理 3.2，存在 $\mathbf{R}^l$ 上的

SG 函数 Φ，使得

$$\|\sin\Theta(Z,W)\| = \Phi(\sigma_1,\cdots,\sigma_l). \tag{4.19}$$

今在 $\mathbf{R}^n$ 上定义一实值函数 Φ^* 如下：对于 $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T\in\mathbf{R}^n$，令

$$\left.\begin{aligned}&\Phi^*(x)=\Phi(|\xi_{i_1}|+|\xi_{i_2}|,\cdots,|\xi_{i_{2l-1}}|+|\xi_{i_{2l}}|),\\&\text{如果}\ |\xi_{i_1}|\geqslant|\xi_{i_2}|\geqslant\cdots\geqslant|\xi_{i_n}|.\end{aligned}\right\}\tag{4.20}$$

为了证明 Φ^* 是 $\mathbf{R}^n$ 上的 SG 函数，显然只需验证定义 3.2 中的性质 III)，即

$$\Phi^*(\boldsymbol{x}+\boldsymbol{y})\leqslant\Phi^*(\boldsymbol{x})+\Phi^*(\boldsymbol{y}),\ \forall\boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n.$$

设 $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T$ 与 $\boldsymbol{y}=(\eta_1,\cdots,\eta_n)^T\in\mathbf{R}^n$，且有

$$|\xi_{i_1}|\geqslant\cdots\geqslant|\xi_{i_n}|,|\eta_{j_1}|\geqslant\cdots\geqslant|\eta_{j_n}|$$

和

$$|\xi_{k_1}+\eta_{k_1}|\geqslant\cdots\geqslant|\xi_{k_n}+\eta_{k_n}|.$$

于是由定义 3.2 的性质 III)，引理 3.4 和引理 3.10，以及 Φ^* 的定义(4.20)，得到

$$\begin{aligned}\Phi^*(\boldsymbol{x}+\boldsymbol{y})&=\Phi(|\xi_{k_1}+\eta_{k_1}|+|\xi_{k_2}+\eta_{k_2}|,\cdots,\\&\qquad|\xi_{k_{2l-1}}+\eta_{k_{2l-1}}|+|\xi_{k_{2l}}+\eta_{k_{2l}}|)\\&\leqslant\Phi(|\xi_{k_1}|+|\eta_{k_1}|+|\xi_{k_2}|+|\eta_{k_2}|,\cdots,\\&\qquad|\xi_{k_{2l-1}}|+|\eta_{k_{2l-1}}|+|\xi_{k_{2l}}|+|\eta_{k_{2l}}|)\\&\leqslant\Phi(|\xi_{k_1}|+|\xi_{k_2}|,\cdots,|\xi_{k_{2l-1}}|+|\xi_{k_{2l}}|)\\&\quad+\Phi(|\eta_{k_1}|+|\eta_{k2}|,\cdots,|\eta_{k_{2l-1}}|+|\xi_{k_{2l}}|)\\&\leqslant\Phi(|\xi_{i_1}|+|\xi_{i_2}|,\cdots,|\xi_{i_{2l-1}}|+|\xi_{i_{2l}}|)\\&\quad+\Phi(|\eta_{j_1}|+|\eta_{j_2}|,\cdots,|\eta_{j_{2l-1}}|+|\eta_{j_{2l}}|)\\&=\Phi^*(\boldsymbol{x})+\Phi^*(\boldsymbol{y}).\end{aligned}$$

因此 Φ^* 是 $\mathbf{R}^n$ 上的 SG 函数．据定理 3.2，若令

$$\|A\|^*=\Phi^*(A),\ \forall A\in\mathbf{C}^{n\times n},$$

则$\|\ \ \|^*$是 $\mathbf{C}^{n\times n}$ 上的酉不变范数．由(4.17)—(4.20)可知

$$\begin{aligned}\|P_Z-P_W\|^*&=\Phi^*(\sigma_1,\sigma_1,\cdots,\sigma_l,\sigma_l,0,\cdots0)\\&=\Phi(2\sigma_1,\cdots,2\sigma_l)=2\Phi(\sigma_1,\cdots,\sigma_l)\\&=2\|\sin\Theta(Z,W)\|,\end{aligned}$$

即

$$\|\sin\Theta(Z,W)\| = \frac{1}{2}\|P_Z - P_W\|^*. \qquad (4.21)$$

因为 $\|P_Z - P_W\|^*$ 是 G_l^n 上的酉不变度量，所以由(4.21)立即推知：$\|\sin\Theta(Z,W)\|$也是 G_l^n 上的酉不变度量. ▭

注 4.1. 当$2l \geqslant n$时，与(4.17)和(4.18)相类似，有关系式

$$\sin\Theta(Z,W) = \mathrm{diag}(\sigma_1,\cdots,\sigma_{n-l},0^{(2l-n)}) \qquad (4.22)$$

和

$$P_Z - P_W = U'\mathrm{diag}(\sigma_1,\sigma_1,\cdots,\sigma_{n-l},\sigma_{n-l},0^{(2l-n)})V', \quad (4.23)$$

其中

$$U' = P\mathrm{diag}\left(\begin{pmatrix}\sigma_1 & \gamma_1\\ -\gamma_1 & \sigma_1\end{pmatrix},\cdots,\begin{pmatrix}\sigma_{n-l} & \gamma_{n-l}\\ -\gamma_{n-l} & \sigma_{n-l}\end{pmatrix},I^{(2l-n)}\right),$$

$$V' = \mathrm{diag}(1,-1,\cdots,1,-1,I^{(2l-n)})P,$$

P 是一排列方阵. 显然 U' 与 $V' \in \mathscr{U}_n$.

由(4.17)、(4.18)、(4.22)和(4.23)可知

$$\|\sin\Theta(Z,W)\|_F = \frac{1}{\sqrt{2}}\|P_Z - P_W\|_F. \qquad (4.24)$$

4.4 其它的度量

从前几段可以看出，能够用不同的办法在 G_l^n 上引进度量. 这里再举两个例子.

第一个例子是陆启铿[29]引进的度量.

对于 G_l^n 上任意二点 Z 与W，令

$$\theta(Z,W) = \arccos\frac{|\det Z^HW|}{\sqrt{\det Z^HZ\det W^HW}}, \quad 0 \leqslant \theta \leqslant \frac{\pi}{2} \qquad (4.25)$$

和

$$\sin\theta(Z,W) = \left\{1 - \frac{|\det Z^HW|^2}{\det Z^HZ\det W^HW}\right\}^{\frac{1}{2}}. \qquad (4.26)$$

下面证明

定理 4.5. 由 (4.25) 和 (4.26) 式所定义的 $\theta(Z,W)$ 与

$\sin\theta(Z,W)$,都是 G_l^n 上的酉不变度量.

证明:

如果 $\theta(Z,W)$是酉不变度量,则容易推知 $\sin\theta(Z,W)$ 亦是酉不变度量. 此外,因为 $\theta(Z,W)$的酉不变性是显然的,所以只需证明 $\theta(Z,W)$是度量.

$\theta(Z,W)$ 显然满足定义 4.1 中的性质 II),以下证明它满足性质 I) 和 III).

首先指出,根据定义(4.25),无妨假定 Z 与W具有第一章定理 3.5 所述的标准形式:

$$Z=\begin{pmatrix} I \\ 0 \\ 0 \end{pmatrix},\quad W=\begin{pmatrix} \Gamma \\ \Sigma \\ 0 \end{pmatrix},\ \text{当 } 2l\leqslant n \text{ 时};$$

$$Z=\begin{pmatrix} I & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix},\quad W=\begin{pmatrix} \Gamma & 0 \\ 0 & I \\ \Sigma & 0 \end{pmatrix},\ \text{当 } 2l\geqslant n \text{ 时}.$$

于是

$$\frac{|\det Z^H W|}{\sqrt{\det Z^H Z \det W^H W}}=\prod_{i=1}^{l}\gamma_i\leqslant 1,$$

其中 $0\leqslant\gamma_1\leqslant\cdots\leqslant\gamma_l\leqslant 1$(当 $2l>n$ 时, $\gamma_{n-l+1}=\cdots=\gamma_l=1$). 因此,

$$\theta(Z,W)=0\Leftrightarrow\prod_{i=1}^{l}\gamma_i=1\Leftrightarrow\gamma_1=\cdots=\gamma_l=1$$

$$\Leftrightarrow W=Z.$$

即 $\theta(Z,W)$ 满足定义 4.1 中的性质 I).

以下证明$\theta(Z,W)$满足三角不等式

$$\theta(Z_1,Z_2)\leqslant\theta(Z_1,Z_0)+\theta(Z_0,Z_2),\ \forall Z_0,Z_1,Z_2\in G_l^n.$$

无妨假定

$$Z_s^H Z_s=I,\ s=0,1,2. \tag{4.27}$$

因此,只要证明

$$\arccos|\det Z_1^H Z_2|$$

$$\leqslant \arccos|\det Z_1^H Z_0| + \arccos|\det Z_0^H Z_2|.$$

用 $\delta_{i_1\cdots i_l}^{(s)}$ 表示由 $Z_s(s=0,1,2)$ 的第 i_1 行，…，第 i_l 行所成方阵的行列式. 根据 Binet-Cauchy 公式(见第一章定理 8.1)，有

$$\det Z_s^H Z_t = \sum_{1\leqslant i_1<\cdots<i_l\leqslant n} \overline{\delta_{i_1\cdots i_l}^{(s)}}\delta_{i_1\cdots i_l}^{(t)},$$

$$s, t = 0, 1, 2. \tag{4.28}$$

把 $\delta_{i_1\cdots i_l}^{(s)}$ 的指标 $i_1\cdots i_l$ 按某一确定的次序排列成一个

$$N = \frac{n!}{l!(n-l)!}$$

维的向量 $\boldsymbol{v}_s \in \mathbb{C}^N$，由(4.27)和(4.28)知，$\boldsymbol{v}_s^H\boldsymbol{v}_s = 1$，$s = 0,1,2$. 于是问题化为求证

$$\arccos|\boldsymbol{v}_1^H\boldsymbol{v}_2| \leqslant \arccos|\boldsymbol{v}_1^H\boldsymbol{v}_0| + \arccos|\boldsymbol{v}_0^H\boldsymbol{v}_2|. \tag{4.29}$$

不难证明，存在 $U \in \mathscr{U}_N$，使得

$\boldsymbol{v}_0 = U(1,0,\cdots,0)^T$,

$\boldsymbol{v}_1 = U(\alpha_1,\alpha_2,0,\cdots,0)^T$, $|\alpha_1|^2 + |\alpha_2|^2 = 1$,

$\boldsymbol{v}_2 = U(\beta_1,\beta_2,\beta_3,0,\cdots,0)^T$, $|\beta_1|^2 + |\beta_2|^2 + |\beta_3|^2 = 1$.

注意到

$$\begin{aligned}|\bar{\alpha}_1\beta_1 + \bar{\alpha}_2\beta_2| &\geqslant |\alpha_1||\beta_1| - |\alpha_2||\beta_2| \\ &= |\alpha_1||\beta_1| - \sqrt{1-|\alpha_1|^2}\sqrt{1-|\beta_1|^2-|\beta_3|^2} \\ &\geqslant |\alpha_1||\beta_1| - \sqrt{1-|\alpha_1|^2}\sqrt{1-|\beta_1|^2} \\ &= \cos(\arccos|\alpha_1| + \arccos|\beta_1|),\end{aligned}$$

由此立即得到

$$\arccos|\bar{\alpha}_1\beta_1 + \bar{\alpha}_2\beta_2| \leqslant \arccos|\alpha_1| + \arccos|\beta_1|,$$

这正是所要证明的不等式(4.29). □

第二个例子是最近 Paige[134] 引进的度量.

对于 $\mathbb{C}_l^n$ 上的任意二点 Z 与 W，令

$$Z_1 = Z(Z^H Z)^{-\frac{1}{2}},\ W_1 = W(W^H W)^{-\frac{1}{2}} \tag{4.30}$$

和

$$d_P(Z,W) = \min_{Q\in\mathscr{U}_l} \|Z_1 - W_1 Q\|_F. \tag{4.31}$$

下面证明

定理 4.6. 由(4.31)式所定义的 $d_P(Z, W)$ 是 $\mathbf{G}_l^n$ 上的酉不变度量.

证明:

$d_P(Z,W)$显然满足定义 4.1—4.2 中的性质 I)、II)与 IV). 以下证明 $d_P(Z,W)$ 满足性质 III), 即满足三角不等式

$$d_P(Z_1,Z_2) \leqslant d_P(Z_1,Z_0) + d_P(Z_0,Z_2),$$
$$\forall Z_0,\ Z_1,\ Z_2 \in \mathbf{C}_l^n. \tag{4.32}$$

无妨假定 $Z_s^H Z_s = I$, $s = 0, 1, 2$, 并且

$$d_P(Z_1,Z_2) = \|Z_1 - Z_2 Q_0\|_F,\ d_P(Z_1,Z_0) = \|Z_0 - Z_1 Q_2\|_F$$

和

$$d_P(Z_0,Z_2) = \|Z_2 - Z_0 Q_1\|_F,$$

其中 $Q_1, Q_2, Q_0 \in \mathscr{U}_l$. 于是由

$$\begin{aligned} \|Z_1 - Z_2 Q_0\|_F &\leqslant \|Z_1 - Z_2 Q_1^H Q_2^H\|_F \\ &= \|Z_1 - Z_0 Q_2^H + Z_0 Q_2^H - Z_2 Q_1^H Q_2^H\|_F \\ &\leqslant \|Z_0 - Z_1 Q_2\|_F + \|Z_2 - Z_0 Q_1\|_F \end{aligned}$$

立即得知不等式(4.32)成立. □

为了今后使用上的方便,特地记

$$d_2(Z,W) = \|\sin\Theta(Z,W)\|_2, \tag{4.33}$$
$$d_L(Z,W) = \sin\theta(Z,W) \tag{4.34}$$

和

$$d_F(Z,W) = \|\sin\Theta(Z,W)\|_F. \tag{4.35}$$

其中 Z 与 $W \in \mathbf{G}_l^n$.

利用(4.30)以及 Z_1 与 W_1 的标准形(见第一章定理 3.5), 可知

$$d_2(Z,W) = \sqrt{1-\gamma_1^2},$$

$$d_L(Z,\ W) = \sqrt{1 - \prod_{i=1}^{l} \gamma_i^2},$$

$$d_F(Z,\ W) = \sqrt{\sum_{i=1}^{l} (1 - \gamma_i^2)}$$

和

$$d_P(Z, W) = \sqrt{2\sum_{i=1}^{l}(1-\gamma_i)},$$

其中 $0 \leqslant \gamma_1 \leqslant \cdots \leqslant \gamma_l \leqslant 1$. 由此得到

$$d_l(Z,W) \leqslant d_L(Z,W) \leqslant d_F(Z,W) \quad \forall Z,W \in \mathbf{G}_l^n \tag{4.36}$$

和

$$d_P(Z,W) \leqslant \sqrt{2}\, d_F(Z,W) \quad \forall Z,\ W \in \mathbf{G}_l^n. \tag{4.37}$$

类似地,可定义 $\mathbf{G}_{l,n}$ 上的 d_2, d_L, d_F 与 d_P.

值得指出的是,当 $l=1, n=2$ 时,$d_l = d_L = d_F$,并且它就是复投影平面 $\mathbf{G}_{1,2}$ 上的弦度量

$$\rho((\alpha,\beta),(\gamma,\delta)) = \frac{|\alpha\delta - \beta\gamma|}{\sqrt{(|\alpha|^2+|\beta|^2)(|\gamma|^2+|\delta|^2)}},$$

$$(\alpha,\beta),(\gamma,\delta) \in \mathbf{G}_{1,2}. \tag{4.38}$$

下面解释弦度量的几何意义. 利用球极投影，可以建立复投影平面 $\mathbf{G}_{1,2}$ 与 Riemann 球面

$$\mathscr{S} = \{(\xi,\eta,\zeta)^T \in \mathbf{R}^3 : \xi^2 + \eta^2 + \zeta^2 = 1\}$$

之间的拓扑、保角对应,这种对应的坐标表示如下: 记点 $(\alpha,\beta) \in \mathbf{G}_{1,2}$ 的非齐次坐标为

$$z = \frac{\alpha}{\beta} = x + iy,\quad x,y \in \mathbf{R},$$

$\mathscr{S}$ 上与点 z 相对应的点 P 的坐标记作(ξ,η,ζ). 其间的关系是

$$(\xi,\eta,\zeta) = \frac{1}{|z|^2+1}(2x, 2y, |z|^2-1)$$

和

$$(x,y) = \frac{1}{1-\zeta}(\xi,\eta).$$

(见图 2-4). 据图 2-4 所示，有

$$\cos 2\phi = \frac{\overrightarrow{OP}\cdot\overrightarrow{OQ}}{\|\overrightarrow{OP}\|\|\overrightarrow{OQ}\|}$$ (·表示点乘,$\|\overrightarrow{OP}\|$ 指向量 $\overrightarrow{OP}$ 的长度)

$$= \frac{4(xu + yv) + (|z|^2 - 1)(|w|^2 - 1)}{(1 + |z|^2)(1 + |w|^2)}$$

$$= 1 - \frac{2|z - w|^2}{(1 + |z|^2)(1 + |w|^2)},$$

由此可得

$$\sin\phi = \frac{|z - w|}{\sqrt{(1 + |z|^2)(1 + |w|^2)}}$$

$$= \rho((\alpha,\beta),(\gamma,\delta)). \tag{4.39}$$

可见，$G_{1,2}$上二点(α,β)与(γ,δ)之间的弦距离等于在球极投影对应下，它们在 Riemann 球面上的对应点之间的弦长之半(当(α,β)与(γ,δ)之中有一点为无穷远点时，可先利用$G_{1,2}$上的一个适当变换，该变换对应着$\mathscr{S}$的一个旋转，使(α,β)与(γ,δ)经该变换后均为有穷点，并且保持弦度量不变，这样便化成了图 2-4 所示的情形).

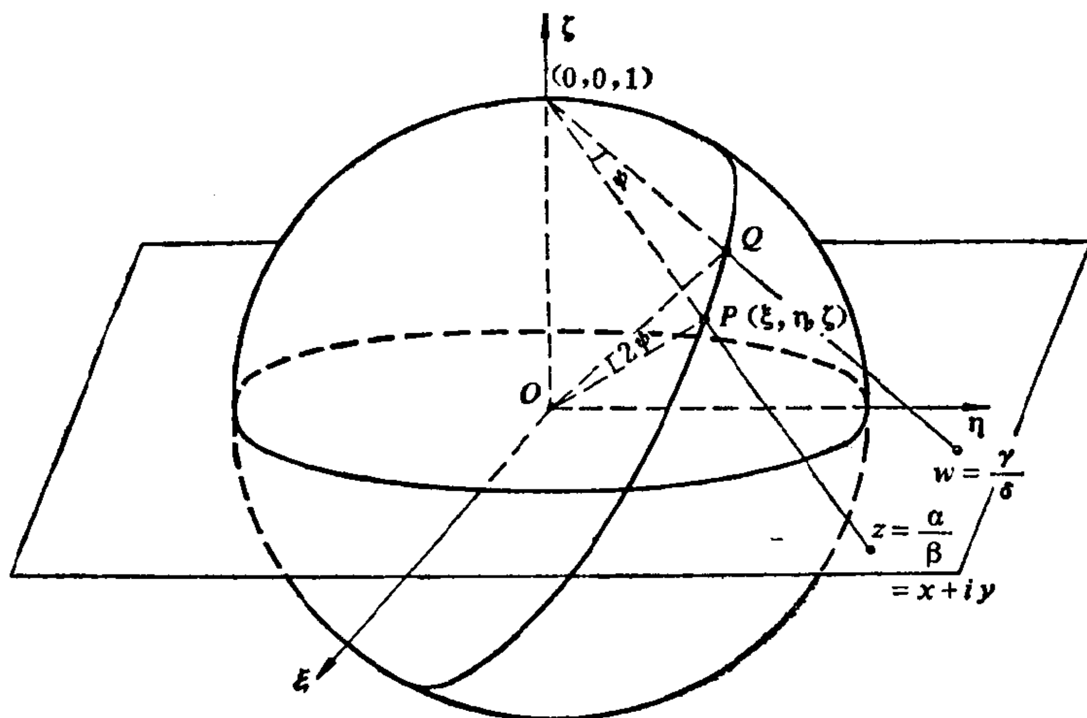


图 2-4

设$(\alpha,\beta),(\gamma,\delta)\in G_{1,2}$，并且$\alpha,\beta,\gamma,\delta\in\mathbf{R}$. 如果把$(\alpha,\beta)$与$(\gamma,\delta)$看作$\mathbf{R}^{1\times 2}$中的点的坐标，并令

$$d((\alpha,\beta),(\gamma,\delta)) = \sqrt{(\alpha-\gamma)^2+(\beta-\delta)^2}, \quad (4.40)$$

则可证

定理 4.7. 设 $\alpha,\beta,\gamma,\delta\in\mathbf{R}$, $\alpha^2+\beta^2=\gamma^2+\delta^2=1$, 并且 $d((\alpha,\beta),(\gamma,\delta))\leqslant\sqrt{2}$, 则有

$$\rho((\alpha,\beta),(\gamma,\delta))\leqslant d((\alpha,\beta),(\gamma,\delta))\leqslant\sqrt{2}\,\rho((\alpha,\beta),(\gamma,\delta)). \quad (4.41)$$

证明:

用 θ 表示 $\mathbf{R}^{1\times 2}$ 中二点 (α,β) 与 (γ,δ) 之间的夹角. 根据假设,有

$$\begin{aligned}\sin\theta &= \sqrt{1-\cos^2\theta} = \sqrt{1-(\alpha\gamma+\beta\delta)^2} = |\alpha\delta-\beta\gamma| \\ &= \rho((\alpha,\beta),(\gamma,\delta)).\end{aligned}$$

另一方面,有

$$d((\alpha,\beta),(\gamma,\delta)) = 2\sin\frac{\theta}{2}.$$

再注意到题设条件等价于 $\theta\in\left[0,\frac{\pi}{2}\right]$; 而当 $\theta\in\left[0,\frac{\pi}{2}\right]$ 时,有

$$\sqrt{2}\sin\frac{\theta}{2}\leqslant\sin\theta\leqslant 2\sin\frac{\theta}{2}.$$

因此不等式(4.41)成立. □

习题

1. 今用 G_1^2 表示 $\mathbf{R}^2$ 内所有 1 维列空间的全体. 试证: $\theta(R(\boldsymbol{x}),R(\boldsymbol{y}))$(见(4.1)式) 与 $\sin\theta(R(\boldsymbol{x}),R(\boldsymbol{y}))$ 都是 G_1^2 上的度量,其中 $\boldsymbol{x},\boldsymbol{y}\in G_1^2$. 又问: $\operatorname{tg}\theta(R(\boldsymbol{x}),R(\boldsymbol{y}))$ 是不是 G_1^2 上的度量?

2. 证明定理 4.1.

3. 设 $Z,W\in\mathbf{C}^{n\times l}(2l\leqslant n)$,满足 $Z^HZ=W^HW=I$. 试利用第一章定理 3.5 给出 $\sin\Theta(Z,W)$(定义见(4.4)式)的酉分解式,并说明 $\sin\Theta(Z,W)$ 的特征值的几何意义.

4. 设 $Z,W\in\mathbf{C}^{n\times l}(2l\leqslant n)$, 满足 $Z^HZ=W^HW=I$. 试证

$$\|\sin\Theta(Z,W)\|_2 = \|(I-P_W)Z\|_2 = \|(I-P_Z)W\|_2$$
$$= \|(I-P_W)P_Z\|_2 = \|(I-P_Z)P_W\|_2.$$

其中 $P_Z = ZZ^H, P_W = WW^H$.

5. 令$\mathscr{Z}$与$\mathscr{W}$表示 $\mathbf{C}^n$ 内的两个 l 维列空间. 令

$$\delta_2(\mathscr{Z},\mathscr{W}) = \max_{\substack{\|z\|_2=1\\ z\in\mathscr{Z}}} \min_{\substack{\|w\|_2=1\\ w\in\mathscr{W}}} \|w-z\|_2.$$

试证

1) $\delta_2(\mathscr{Z},\mathscr{W}) = \sqrt{2(1-\sqrt{1-\gamma_2^2(\mathscr{Z},\mathscr{W})})}$,

其中 $\gamma_2(\mathscr{Z},\mathscr{W})$ 表示借助于向量的 Euclid 范数 $\|\ \|_2$ 所定义的 $\gamma(\mathscr{Z},\mathscr{W})$(见(4.7)式);

2) $\delta_2(\mathscr{Z},\mathscr{W})$是 $\mathbf{G}_l^n$ 上的酉不变度量;

3) $\gamma_2(\mathscr{Z},\mathscr{W}) \leqslant \delta_2(\mathscr{Z},\mathscr{W}) \leqslant 2\gamma_2(\mathscr{Z},\mathscr{W})$.

第二章说明

在数值分析与矩阵论中,矩阵范数已经成了十分重要的概念,这首先应该归功于 Householder 和 Bauer 等人的工作,他们从五十年代末开始,致力于矩阵范数及其在数值分析中的应用的研究(见[105]—[109]和[43]—[51]).

关于矩阵范数,本章主要讨论算子范数和酉不变范数,它们都是相容范数,在矩阵扰动分析中应用最多. 除了算子范数和酉不变范数,还有更广意义下的矩阵范数;对此有兴趣的读者,可以参看 Deutsch 的论文[78]—[81].

投影空间上的度量(包括子空间之间的距离),在矩阵扰动分析中显得愈来愈重要(参看[13]、[61]、[75]和[155]),可是至今尚未见到系统的论述;作者把近几年的工作整理编写了本章 §4,主要参考[7]、[16]、[18]、[75]、[117]、[155]和[184]. 有些问题,例如复投影空间上的酉不变度量,尚待进一步研究.

第三章　特征值问题扰动分析

§1　特征值问题的稳定性

设 $A=(\alpha_{ij})\in\mathbf{C}^{n\times n}$．研究特征值问题 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ ($\boldsymbol{x}\in\mathbf{C}^n$, $\boldsymbol{x}\neq 0$, $\lambda\in\mathbf{C}$) 的稳定性，就是研究特征值 λ 及其相应的特征向量 $\boldsymbol{x}$ 是否连续地依赖于 A 的元素 $\alpha_{ij}(1\leqslant i,j\leqslant n)$．显然，这里所指的稳定性，是特征值问题 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 本身的性质(即矩阵 A 本身的性质)，并不依赖于算法．但另一方面，在数值计算时，由于实际问题中的数据往往带有误差，在计算机上表示矩阵的元素也会有误差．对于算法进行误差分析，结果表明，通过计算得到的特征值与特征向量，往往是经过扰动之后的特征值问题 $\tilde{A}\tilde{\boldsymbol{x}}=\tilde{\lambda}\tilde{\boldsymbol{x}}$ 的特征值 $\tilde{\lambda}$ 与特征向量 $\tilde{\boldsymbol{x}}$，其中 $\tilde{A}=A+E$，而对于 $E=(\varepsilon_{ij})\in\mathbf{C}^{n\times n}$，只知道 $|\varepsilon_{ij}|(1\leqslant i,j\leqslant n)$ 的某个上界或者 $\|E\|$ 的某个上界；因此，研究特征值问题的稳定性，以及研究如何利用 $|\varepsilon_{ij}|(1\leqslant i,j\leqslant n)$ 的上界或者 $\|E\|$ 的上界，尽可能精确地估计 $\tilde{\lambda}$ 与 λ 之间，以及 $\tilde{\boldsymbol{x}}$ 与 $\boldsymbol{x}$ 之间的差距，在特征值问题的数值计算中，有着重要的意义．

首先要指出的是，矩阵的特征向量不一定是稳定的．例如

$$A_\varepsilon=\begin{pmatrix}1 & 0 & 0\\ 0 & 1+\varepsilon & 1\\ 0 & 0 & 1+\varepsilon\end{pmatrix}$$

的特征向量在 $\varepsilon=0$ 处不连续．但矩阵的特征值是稳定的，这一点早已为下面的 Ostrowski 定理所阐明．

1.1　特征值的连续性

定理 1.1 (Ostrowski [130])，设 $A=(\alpha_{ij})$，$B=(\beta_{ij})\in\mathbf{C}^{n\times n}$，$\lambda(A)=\{\lambda_i\}$，$\lambda(B)=\{\mu_i\}$．令

$$m_\alpha = \max_{i,j}(|\alpha_{ij}|, |\beta_{ij}|), \tag{1.1}$$

$$\|B - A\|_\alpha = \frac{1}{n}\sum_{i,j=1}^{n} |\beta_{ij} - \alpha_{ij}| \tag{1.2}$$

和

$$\delta = (n+2)m_\alpha^{1-\frac{1}{n}} \|B - A\|_\alpha^{1/n}. \tag{1.3}$$

则对于任一 $\mu \in \lambda(B)$，必有 $\lambda_i(\mu) \in \lambda(A)$，使得

$$|\lambda_i(\mu) - \mu| < \delta; \tag{1.4}$$

并且存在 $1,\cdots,n$ 的一个适当排列 $\pi(1),\cdots,\pi(n)$，使得

$$|\lambda_i - \mu_{\pi(i)}| < (2n-1)\delta, i = 1,2,\cdots n. \tag{1.5}$$

注 1.1. 从定理 1.1 容易看出，当 $B \to A$ 时，存在一个固定的排列 π，使得 $\mu_{\pi(i)} \to \lambda_i (i = 1,\cdots,n)$.

先证明两个引理.

引理 1.1. 在定理 1.1 的假设条件下，如果令

$$\varphi(\lambda) = \det(\lambda I - A), \quad \psi(\lambda) = \det(\lambda I - B),$$

则对于任一满足

$$|\lambda| \leqslant n m_\alpha \tag{1.6}$$

的 λ，必有

$$|\varphi(\lambda) - \psi(\lambda)| < \delta^n. \tag{1.7}$$

证明:

首先把 $\{\alpha_{ij}\}$ 和 $\{\beta_{ij}\}$ 按同样的顺序排列为 $\alpha_1, \alpha_2 \cdots, \alpha_{n^2}$ 和 $\beta_1, \beta_2, \cdots, \beta_{n^2}$，则存在一个多项式 P，使得

$$\varphi(\lambda) = P(\alpha_1,\cdots,\alpha_{n^2}), \quad \psi(\lambda) = P(\beta_1,\cdots,\beta_{n^2}).$$

于是

$$\varphi(\lambda) - \psi(\lambda) = \sum_{k=1}^{n^2} \Delta_k,$$

其中

$$\Delta_k = P(\alpha_1,\cdots,\alpha_{k-1},\alpha_k,\beta_{k+1},\cdots,\beta_{n^2}) - P(\alpha_1,\cdots,\alpha_{k-1},\beta_k,\beta_{k+1},\cdots,\beta_{n^2}).$$

根据行列式展开可知

$$\Delta_k = \pm(\alpha_k - \beta_k)\ \det T_k \quad (k = 1,\cdots,n^2),$$

其中 $T_k \in \mathbf{C}^{(n-1)\times(n-1)}$，它的每一行至多有一个 $\lambda - \alpha_i$ 或 $\lambda - \beta_j$，其它的元素为 α_m 与 β_l. 所以 T_k 的每一行的 Euclid 范数不超过 $\sqrt{(n-2)m_\alpha^2 + (n+1)^2 m_\alpha^2} < (n+2)m_\alpha$. 再据 Hadamard 不等式，有

$$|\det T_k| \leqslant [(n+2)m_\alpha]^{n-1}.$$

因此，

$$\begin{aligned}|\varphi(\lambda) - \psi(\lambda)| &\leqslant (n+2)^{n-1} m_\alpha^{n-1} \sum_{k=1}^{n^2} |\alpha_k - \beta_k| \\ &< (n+2)^n m_\alpha^{n-1} \cdot \frac{1}{n} \sum_{k=1}^{n^2} |\alpha_k - \beta_k| \\ &= (n+2)^n m_\alpha^{n-1} \|B - A\|_\alpha = \delta^n.\end{aligned}$$

□

引理 1.2. 设 $A, B \in \mathbf{C}^{n\times n}$. 令

$$\chi_t(z) = \det\{zI^{(n)} - [(1-t)A + tB]\},\ 0 \leqslant t \leqslant 1. \quad (1.8)$$

又设 $\mathscr{D}$ 是 $\mathbf{C}$ 内一闭域，其边界 $\partial\mathscr{D}$ 由有穷段光滑曲线组成，并且

$$\chi_t(z) \neq 0,\ \forall t \in [0,1],\ \forall z \in \partial\mathscr{D},$$

则 $\chi_t(z)$ 在 $\mathscr{D}$ 内的零点个数 $N(t)$ 与 t 在$[0,1]$上的取值无关.

证明:

任取 $t_0 \in [0,1]$，令

$$u(z) = \chi_{t_0}(z)$$

及

$$p = \min_{z\in\partial\mathscr{D}} |\chi_{t_0}(z)| > 0.$$

此外，对于 $t_0 + \varepsilon \in [0,1]$，令

$$v(z) = \chi_{t_0+\varepsilon}(z) - \chi_{t_0}(z).$$

因为 $\chi_t(z)$ 是闭集 $[0,1] \times \partial\mathscr{D}$ 上的连续函数，所以当 $|\varepsilon|$ 足够小时，可以使得

$$\max_{z\in\partial\mathscr{D}} |v(z)| < p.$$

于是 $\mathscr{D}$ 上的两个单值解析函数 $u(z)$ 和 $v(z)$ 满足

$$|u(z)| > |v(z)|, \quad \forall z \in \partial \mathscr{D}.$$

据 Rouché 定理*, $u(z) + v(z) = \chi_{t_0+\varepsilon}(z)$ 与 $u(z) = \chi_{t_0}(z)$ 在 $\mathscr{D}$ 内有相同个数的零点. 因此，在 t_0 的 $|\varepsilon|$ 邻域内，$\chi_t(z)$ 在 $\mathscr{D}$ 内的零点个数 $N(t)$ 是个常数;从而 $N(t)$ 是 $t \in [0,1]$ 的连续函数. 但是 $N(t)$ 只能是整数,所以 $N(t)$ 在[0,1]上是常数.

□

定理 1.1 的证明:

1) 设 μ 是 B 的任一特征值. 根据(1,1)及第二章定理 2.4,有

$$|\mu| \leqslant \|B\|_\infty = \max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} |\beta_{ij}| \leqslant n m_\alpha.$$

因此，利用引理 1.1 得到

$$\prod_{i=1}^{n} |\mu - \lambda_i| = |\varphi(\mu)| = |\varphi(\mu) - \psi(\mu)| < \delta^n.$$

所以至少有一个 $\lambda_i(\mu)$, 适合

$$|\lambda_i(\mu) - \mu| < \delta,$$

这就是不等式(1.4).

2) 如果 $|\lambda_i - \lambda_j| < 2\delta$, 则称 λ_i 与 λ_j 相邻; 如果存在 $\lambda_{\nu_1}, \cdots, \lambda_{\nu_k}$, 使得点列 $\lambda_i, \lambda_{\nu_1}, \cdots, \lambda_{\nu_k}, \lambda_j$ 中,除 λ_i 外,每一点都与前面的一点相邻，则称 λ_i 与 λ_j 相连. 现将点集 $\{\lambda_i\}$ 划分成若干个子集 $\Omega_1, \cdots, \Omega_s$, 划分的原则是: 所有相连的点属于同一个子集,不相连的点属于不同的子集.

令

$$\mathscr{D}(\lambda, r) = \{z \in \mathbf{C}: |z - \lambda| \leqslant r\}, \lambda \in \mathbf{C}, r > 0 \tag{1.9}$$

和

$$G_k = \bigcup_{\lambda_i \in \Omega_k} \mathscr{D}(\lambda_i, \delta), \quad k = 1, \cdots, s. \tag{1.10}$$

---

* **Rouché 定理** 设 $\mathscr{D}$ 是 $\mathbf{C}$ 内一闭域,其边界 $\partial\mathscr{D}$ 由有穷段光滑曲线组成. 如果 $u(z)$ 与 $v(z)$ 在 $\mathscr{D}$ 上单值解析,并且

$$|u(z)| > |v(z)|, \quad \forall z \in \partial\mathscr{D},$$

则 $u(z) + v(z)$ 与 $u(z)$ 在 $\mathscr{D}$ 内有相同个数的零点.

G_k 的边界记作 ∂G_k，它由有限个圆弧构成.

现在证明：(1.8) 所示的函数 $\chi_t(z)$ $(0 \leqslant t \leqslant 1)$ 在 $\bigcup_{k=1}^{s} \partial G_k$ 上无零点. 这是因为，如果对某一个 $t \in [0, 1]$，存在 $\lambda' \in \mathbf{C}$，使得 $\chi_t(\lambda') = 0$，则有

$$\prod_{i=1}^{n} |\lambda' - \lambda_i| = |\varphi(\lambda')| = |\varphi(\lambda') - \chi_t(\lambda')|.$$

其中

$$\begin{aligned} \chi_t(\lambda') &= \det\{\lambda' I - [(1-t)A + tB]\} \\ &= \det\{\lambda' I - \tilde{B}\} \equiv \tilde{\psi}(\lambda'), \\ \tilde{B} &= (1-t)A + tB \equiv (\tilde{\beta}_{ij}), \end{aligned}$$

每个 $\tilde{\beta}_{ij}$ 适合 $|\tilde{\beta}_{ij}| \leqslant m_a$，从而 $|\lambda'| \leqslant nm_a$. 于是，根据引理 1.1，由

$$\prod_{i=1}^{n} |\lambda' - \lambda_i| = |\varphi(\lambda') - \tilde{\psi}(\lambda')|$$

得出

$$\prod_{i=1}^{n} |\lambda' - \lambda_i| < \delta^n.$$

因此必存在某一个 $\lambda_{i'}$，使得 $|\lambda' - \lambda_{i'}| < \delta$，即 λ' 必位于某一个 G_k 内. 所以 $\chi_t(z)(0 \leqslant t \leqslant 1)$ 在 $\bigcup_{k=1}^{s} \partial G_k$ 上无零点.

再由引理 1.2，$\chi_t(z)$ 在每个 G_k 内的零点个数 $N_k(t)$ $(k = 1, \cdots, s)$ 与 t 在$[0,1]$上的取值无关. 且注意到 $\chi_0(\lambda) = \varphi(\lambda)$，$\chi_1(\lambda) = \psi(\lambda)$，因此有，$\varphi(\lambda)$与 $\psi(\lambda)$ 在每个 $G_k(k = 1, \cdots, s)$ 内有相同个数的根. 比如在 G_1 内，$\varphi(\lambda)$ 与 $\psi(\lambda)$ 各有 N_1 个根. 根据 G_k 的构造可知，在 G_1 内，$\varphi(\lambda)$ 的每个根 λ_i 与 $\psi(\lambda)$ 的每个根 μ_j 之间的距离，适合

$$|\lambda_i - \mu_j| \leqslant (2N_1 - 1)\delta \leqslant (2n - 1)\delta.$$

同理考虑其它的集合 G_k ($k=2,\cdots,s$). 所以存在 $\{\mu_i\}$ 的一个适当的排列 $\mu_{\pi(1)},\cdots,\mu_{\pi(n)}$, 使得

$$|\lambda_i-\mu_{\pi(i)}|\leqslant(2n-1)\delta, i=1,\cdots,n. \qquad \square$$

注 1.2 定理 1.1 的估计式(1.5)虽然看来有些粗糙,但它揭示了矩阵特征值连续依赖于矩阵元素这一重要性质; 并且如果不对矩阵 A 或 B 附加任何限制,则估计式(1.5)右端 $\|B-A\|_\alpha^{\frac{1}{n}}$ 的指数 $\frac{1}{n}$ 已不能再改进, 例如考虑下列 $n\times n$ 矩阵 A 与 B:

$$A=\begin{pmatrix} 1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 1 \end{pmatrix},\quad B=\begin{pmatrix} 1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \varepsilon & & & 1 \end{pmatrix},\ 0<\varepsilon\ll 1,$$

有

$$\|B-A\|_\alpha=\frac{\varepsilon}{n}$$

和

$$\lambda(A)=\{1,\cdots,1\},\quad \lambda(B)=\{\mu_1,\cdots,\mu_n\},$$

其中

$$|\mu_i-1|=\varepsilon^{\frac{1}{n}}=n^{\frac{1}{n}}\|B-A\|_\alpha^{\frac{1}{n}}, i=1,\cdots,n.$$

1.2 扰动性质的数学描述

设 $A,B\in\mathbb{C}^{n\times n}$. $\lambda(A)=\{\lambda_i\}$, $\lambda(B)=\{\mu_i\}$. 用 $d(\lambda(A),\lambda(B))$ 表示 $\lambda(A)$ 与 $\lambda(B)$ 之间的某种差距, $D(A,B)$ 表示 A 与 B 之间的某种差距. 关于 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 的特征值的扰动性质, 在数学上一般用

$$d(\lambda(A),\lambda(B))\leqslant F(D(A,B))$$

描述,有时也用

$$f(d(\lambda(A),\lambda(B)))\leqslant D(A,B)$$

描述. 其中 F 与 f 是比较简单的函数.

$D(A,B)$ 往往取范数 $\|B-A\|_\nu$, 此处 ν 可以是 1, 2, ∞,

F, α 等等.

关于 $d(\lambda(A), \lambda(B))$, 通常考虑下列三种形式:

1) B 对于 A 的谱改变量

$$s_A(B) \equiv \max_{1 \leqslant j \leqslant n} \{ \min_{1 \leqslant i \leqslant n} |\lambda_i - \mu_j| \}; \tag{1.11}$$

2) A 与 B 的特征值改变量

$$v(A,B) \equiv \min_{\pi} \{ \max_{1 \leqslant i \leqslant n} |\lambda_i - \mu_{\pi(i)}| \}, \tag{1.12}$$

其中 π 是 $1, \cdots, n$ 的某个排列,上式右端的 $\min\limits_{\pi}$ 表示在 $1, \cdots, n$ 的所有可能的排列上取最小值;

3) A 与 B 的谱的 Euclid 距离

$$e(A,B) \equiv \min_{\pi} \sqrt{\sum_{i=1}^{n} |\lambda_i - \mu_{\pi(i)}|^2}, \tag{1.13}$$

其中 π 是 $1, \cdots, n$ 的某个排列,上式右端的 $\min\limits_{\pi}$ 表示在 $1, \cdots, n$ 的所有可能的排列上取最小值.

A 与 B 的对应特征向量 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 之间的差距,一般用夹角 $\theta(R(\boldsymbol{x}), R(\boldsymbol{y}))$ (见第二章 § 4(4.1)) 或者它的正弦 $\sin\theta(R(\boldsymbol{x}), R(\boldsymbol{y}))$ 表示.

$s_A(B) \leqslant \gamma$ 的几何意义: 如果令

$$\mathscr{D}_i = \{z \in \mathbf{C} : |z - \lambda_i| \leqslant \gamma\}, \quad i = 1, \cdots, n,$$

则 $s_A(B) \leqslant \gamma$ 表示 $\lambda(B) \subseteq \bigcup\limits_{i=1}^{n} \mathscr{D}_i$.

$v(A,B) \leqslant \gamma$ 的几何意义: $v(A,B) \leqslant \gamma$ 表示存在 $\{\lambda_i\}$ 与 $\{\mu_i\}$ 的适当排列 $\lambda_1, \cdots, \lambda_n$ 与 $\mu_{k_1}, \cdots, \mu_{k_n}$, 使得

$$|\lambda_i - \mu_{k_i}| \leqslant \gamma, \quad i = 1, \cdots, n.$$

$e(A,B) \leqslant \gamma$ 的几何意义. $e(A,B) \leqslant \gamma$ 表示存在 $\{\lambda_i\}$ 与 $\{\mu_i\}$ 的适当排列 $\lambda_1, \cdots, \lambda_n$ 与 $\mu_{l_1}, \cdots, \mu_{l_n}$, 如果令 $\boldsymbol{a} = (\lambda_1, \cdots, \lambda_n)^T$ 和 $\boldsymbol{b} = (\mu_{l_1}, \cdots, \mu_{l_n})^T$, 则

$$\|\boldsymbol{a} - \boldsymbol{b}\|_2 \leqslant \gamma.$$

利用上述记号,定理 1.1 的结论(1.4)和(1.5)可简单地记作

$$s_A(B) < \delta, \quad v(A,B) < (2n-1)\delta.$$

容易证明(本节习题1)

$$s_A(B) \leqslant v(A,B) \leqslant c(A,B); \tag{1.14}$$

另一方面,有下述结论(参考[33]和[86]):

定理 1.2. 设 $A, B \in \mathbf{C}^{n\times n}$. 令

$$h_A(B) = \max_{0\leqslant t\leqslant 1} s_A((1-t)A + tB). \tag{1.15}$$

则有

$$v(A,B) \leqslant (2n-1)h_A(B). \tag{1.16}$$

证明:

由 $h_A(B)$ 的定义可知, $(1-t)A+tB(t\in[0,1])$ 的特征值必包含在

$$G = \bigcup_{i=1}^{n} \mathscr{D}(\lambda_i, h_A(B)) \tag{1.17}$$

之中, 其中 $\mathscr{D}(\lambda, r)$ 的定义见 (1.9). 设G包含k个连通分支$G_1,\cdots,G_k$, 每个 G_j 中包含有A的 $N_j(j=1,\cdots,k)$ 个特征值. 根据定理1.1, $(1-t)A+tB$ 的特征值连续地依赖于t,因此$(1-t)A+tB$ $(0<t\leqslant 1)$ 在每个 G_j 中亦恰有 N_j 个特征值(否则, 对某一个 $t\in[0,1]$, $(1-t)A+tB$ 必有特征值位于$\mathbf{C}\backslash G$ 之中,这与 $h_A(B)$ 及G的定义相矛盾); 特别地,取 $t=1$,可知B与A在每个 G_j 中都有 N_j个特征值. 所以存在 $1,\cdots,n$ 的一个适当的排列 $\pi(1),\cdots,\pi(n)$, 使得

$$|\lambda_i - \mu_{\pi(i)}| \leqslant (2N-1)h_A(B), \quad i=1,\cdots,n, \tag{1.18}$$

其中 $N=\max\limits_{1\leqslant k\leqslant s} N_k \leqslant n$. 由(1.18)立即得到(1.16). □

推论 1.1. 设 $A,B\in\mathbf{C}^{n\times n}$. $h_A(B)$ 如(1.15)所示,并且满足

$$h_A(B) \leqslant F(\nu(B-A)), \tag{1.19}$$

其中 ν 是 $\mathbf{C}^{n\times n}$ 上的一个非负实值函数, F 是 $\mathbb{R}$ 上的一个非负实值函数. 又设A有 l 个不同的特征值. 则

$$v(A,B) \leqslant (2l-1)F(\nu(B-A)). \tag{1.20}$$

证明:

考虑 (1.17) 所示的点集 G. 只需注意到一点，即 G 的每个连通分支 G_j 都是不超过 l 个圆盘 $\mathscr{D}(\lambda_i, h_A(B))$ 的并集. 因此 (1.18) 式可改写为

$$|\lambda_i - \mu_{\pi(i)}| \leqslant (2l-1)h_A(B), \quad i = 1, \cdots, n.$$

再联系到(1.19)，即得到(1.20). □

定理 1.2 和推论 1.1 启发我们，在许多情况下，可以首先着眼于求 $s_A(B)$ 的上界.

习题

1. 证明

$$s_A(B) \leqslant v(A, B) \leqslant e(A, B).$$

2. 设

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1-10^{-4} \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ -10^{-2} & 1-10^{-4} \end{pmatrix}.$$

计算 $s_A(B)$，并应用 Ostrowski 定理给出 $s_A(B)$ 的上界.

3. 设 λ 是 A 的单特征值，$\boldsymbol{x}$ 与 $\boldsymbol{y}^H$ 分别是 A 属于 λ 的右特征向量与左特征向量(即 $A\boldsymbol{x} = \lambda\boldsymbol{x}, \boldsymbol{y}^H A = \lambda\boldsymbol{y}^H$). 试证 $\boldsymbol{y}^H\boldsymbol{x} \neq 0$.

§2 Gerschgorin 理论

2.1 Gerschgorin 定理

首先证明一个排除定理（即对于矩阵 A，在 $\mathbf{C}$ 中划定一个区域 $\mathscr{D}_A$，证明 $\mathbf{C}\backslash\mathscr{D}_A$ 中无 A 的特征值).

定理 2.1. 设 $A \in \mathbf{C}^{n\times n}$. 任取一矩阵 $B \in \mathbf{C}^{n\times n}$，令

$$\mathscr{D}_A = \{z \in \mathbf{C}: \|(zI-B)^{-1}(A-B)\| \geqslant 1 \text{ 或 } \det(zI-B) = 0\}, \tag{2.1}$$

其中$\| \ \|$是任一相容的矩阵范数. 则

$$\lambda(A) \subseteq \mathscr{D}_A. \tag{2.2}$$

证明:

设 λ 是 A 的任一特征值，$\boldsymbol{x}$ 是 A 属于 λ 的一个特征向量:

$Ax = \lambda x$. 于是

$$(A - B)x = (\lambda I - B)x.$$

若 $\lambda \in \lambda(B)$，则显然有 $\lambda \in \mathscr{D}_A$. 若 $\lambda \notin \lambda(B)$，则 $\lambda I - B$ 非奇异，从而

$$x = (\lambda I - B)^{-1}(A - B)x.$$

据第二章定理 2.3，对于相容的矩阵范数 $\| \ \|$，必存在与之相容的向量范数 ν. 因此

$$\nu(x) \leqslant \|(\lambda I - B)^{-1}(A - B)\|\nu(x),$$

即 $\lambda \in \mathscr{D}_A$. □

定理 2.2. (Gerschgorin 定理). 设 $A = (\alpha_{ij}) \in \mathbb{C}^{n\times n}$. 令 $\boldsymbol{\alpha}_i' = (\alpha_{i1}, \cdots, \alpha_{i,i-1}, \alpha_{i,i+1}, \cdots, \alpha_{in})^T$ 和

$$G_i(A) = \{z \in \mathbb{C}: |z - \alpha_{ii}| \leqslant \|\boldsymbol{\alpha}_i'\|_1\}, \quad i = 1, \cdots, n. \tag{2.3}$$

则

$$1)\ \lambda(A) \subset \bigcup_{i=1}^{n} G_i(A); \tag{2.4}$$

2) 如果在(2.3)所示的圆盘中，有 m 个互相连通，且与其余的 $n - m$ 个不连通，则在此 m 个圆盘所成的连通区域中，恰有 A 的 m 个特征值.

证明:

1) 在定理 2.1 中，取 $B = \operatorname{diag}(\alpha_{11}, \cdots, \alpha_{nn})$ 和 $\| \ \| = \| \ \|_\infty$，则

$$\lambda(A) \subseteq \mathscr{D}_A = \left\{z \in \mathbb{C}: \max_{1\leqslant i\leqslant n} \frac{\|\boldsymbol{\alpha}_i'\|_1}{|z - \alpha_{ii}|} \geqslant 1 \quad 或 \quad \prod_{i=1}^{n} (z - \alpha_{ii}) = 0\right\}.$$

容易看出

$$\mathscr{D}_A = \bigcup_{i=1}^{n} G_i(A).$$

2) 记 $D = \operatorname{diag}(\alpha_{11}, \cdots, \alpha_{nn})$ 和

$$A = D + C, \quad A_t = D + tC, \quad 0 \leqslant t \leqslant 1,$$

其中 C 是由 A 的对角元素置零得到的矩阵.

不失一般性，假设(2.3)中的前 m 个圆盘构成一连通区域，并与后 $n-m$ 个圆盘分离. 由 1) 知,若令

$$G_i(A_t)=\{z\in\mathbb{C}: |z-\alpha_{ii}|\leqslant t\|\alpha_i'\|_1\},\quad i=1,\cdots,n,$$

则

$$\lambda(A_t)\subset\bigcup_{i=1}^{n}G_i(A_t);$$

并且有

$$G_i(A_t)\subseteq G_i(A),\ G_i(A_0)=G_i(D)=\alpha_{ii},\ G_i(A_1)=G_i(A),\quad i=1,\cdots,n. \tag{2.5}$$

记 $\lambda(A)=\{\lambda_i\}$, $\lambda(A_t)=\{\lambda_i(t)\}$. 根据定理 1.1，每个 $\lambda_i(t)$ 连续地依赖于 $t\in[0,1]$,因此当 t 从 0 变化到 1 时,每个 $\lambda_i(t)$ 表示了 $\mathbb{C}$ 中的一条连续曲线，它的始点为 $\lambda_i(0)=\alpha_{ii}$，终点为 $\lambda_i(1)=\lambda_i$, $i=1,\cdots,n$. 由(2.5)知,当 t 在 $[0,1]$ 上连续变化时，A_t 的前 m 个圆盘始终与后 $n-m$ 个圆盘分离，因而，以 $\alpha_{11},\cdots,\alpha_{mm}$ 为始点的 m 条连续曲线 $\lambda_1(t),\cdots,\lambda_m(t)$ 全部位于 $\bigcup_{i=1}^{m}G_i(A)$ 之中,所以它们的终点 $\lambda_1,\cdots,\lambda_m\in\bigcup_{i=1}^{m}G_i(A)$. 同理可证, $\bigcup_{i=1}^{m}G_i(A)$ 内不含有其它的曲线 $\lambda_j(t)$ $(m+1\leqslant j\leqslant n)$, 因而不含有 $\lambda_{m+1},\cdots,\lambda_n$. □

例 2.1.

$$A=\begin{pmatrix}2&1&0\\ \frac{1}{4}&3&\frac{1}{4}\\ \frac{1}{2}&\frac{1}{2}&6\end{pmatrix}.$$

A 的 Gerschgorin 圆盘为

$$G_1(A)=\{z\in\mathbb{C}: |z-2|\leqslant 1\},$$

$$G_2(A)=\left\{z\in\mathbb{C}:|z-3|\leqslant\frac{1}{2}\right\},$$

$$G_3(A)=\{z\in\mathbb{C}:|z-6|\leqslant 1\}$$

(见图 3-1).

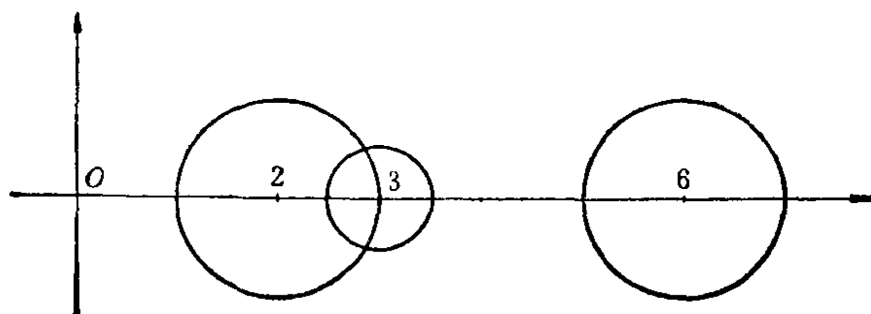


图 3-1

A 在 $G_1(A)\cup G_2(A)$ 中有 2 个特征值，在 $G_3(A)$ 中有 1 个特征值.

50 多年来，Gerschgorin 理论有了很大的发展；可参阅文献 [38],[88],[89],[113],[124],[125]和[179]—[181].

下面仅举例介绍 Wilkinson 如何运用 Gerschgorin 定理去估计特征值的扰动界限(详见[189]).

2.2 应用举例

设 $A\in\mathbb{C}^{n\times n}$ 是一可正规化阵，λ_1 是它的 $p(\geqslant 1)$ 重特征值. $B=(\beta_{ij})\in\mathbb{C}^{n\times n}$，其中 $|\beta_{ij}|\leqslant 1(1\leqslant i,j\leqslant n)$. 现考察 $\lambda(A+\varepsilon B)$ 与 λ_1 的关系，此处 $0<\varepsilon\ll 1$.

据假设，存在非奇异阵 $X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)\in\mathbb{C}^{n\times n}$ 和 $Z=(\boldsymbol{z}_1,\cdots,\boldsymbol{z}_n)$，$Z^H=X^{-1}$，使得

$$Z^HAX=\mathrm{diag}(\lambda_1,\cdots,\lambda_1,\lambda_{p+1},\cdots,\lambda_n),\qquad(2.6)$$

其中 $\|\boldsymbol{x}_i\|_2=1,i=1,\cdots,n$. 令 $\boldsymbol{y}_i=\boldsymbol{z}_i/\|\boldsymbol{z}_i\|_2$ 和

$$s_i=\boldsymbol{y}_i^H\boldsymbol{x}_i,i=1,\cdots,n.\qquad(2.7)$$

显然 $\boldsymbol{x}_i$ 与 $\boldsymbol{y}_i^H$ 分别是 A 的单位右特征向量和单位左特征向量，并且有

$$
\left.\begin{array}{l}
s_i = \dfrac{z_i^H x_i}{\|z_i\|_2} = \dfrac{1}{\|z_i\|_2},\ 0 < s_i \leqslant 1, \\
y_i = s_i z_i,\ i = 1, \cdots, n.
\end{array}\right\} \tag{2.8}
$$

于是

$$
Z^H(A + \varepsilon B)X = \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_1 & & \\ & & & \lambda_{p+1} & \\ & & & & \ddots \\ & & & & & \lambda_n \end{pmatrix} + \varepsilon \begin{pmatrix} \dfrac{\gamma_{11}}{s_1} & \cdots & \dfrac{\gamma_{1n}}{s_1} \\ \vdots & & \vdots \\ \dfrac{\gamma_{n1}}{s_n} & \cdots & \dfrac{\gamma_{nn}}{s_n} \end{pmatrix},
$$

其中 $\gamma_{ij} = y_i^H B x_j$，满足

$$
|\gamma_{ij}| \leqslant n. \tag{2.9}
$$

现取一对角阵

$$
D = \operatorname{diag}\Big(\underbrace{\frac{\sigma}{\varepsilon}, \cdots, \frac{\sigma}{\varepsilon}}_{p\text{个}}, 1, \cdots, 1\Big),
$$

其中 σ 为一待定正数. 作相似变换

$$
D^{-1}Z^H(A + \varepsilon B)XD = \operatorname{diag}(\lambda_1, \cdots, \lambda_1, \lambda_{p+1}, \cdots, \lambda_n)
$$

$$
+ \left(\begin{array}{ccc|ccc}
\dfrac{\varepsilon\gamma_{11}}{s_1} & \cdots & \dfrac{\varepsilon\gamma_{1p}}{s_1} & \dfrac{\varepsilon^2\gamma_{1,p+1}}{\sigma s_1} & \cdots & \dfrac{\varepsilon^2\gamma_{1n}}{\sigma s_1} \\
\vdots & & \vdots & \vdots & & \vdots \\
\dfrac{\varepsilon\gamma_{p1}}{s_p} & \cdots & \dfrac{\varepsilon\gamma_{pp}}{s_p} & \dfrac{\varepsilon^2\gamma_{p,p+1}}{\sigma s_p} & \cdots & \dfrac{\varepsilon^2\gamma_{pn}}{\sigma s_p} \\
\hline
\dfrac{\sigma\gamma_{p+1,1}}{s_{p+1}} & \cdots & \dfrac{\sigma\gamma_{p+1,p}}{s_{p+1}} & \dfrac{\varepsilon\gamma_{p+1,p+1}}{s_{p+1}} & \cdots & \dfrac{\varepsilon\gamma_{p+1,n}}{s_{p+1}} \\
\vdots & & \vdots & \vdots & & \vdots \\
\dfrac{\sigma\gamma_{n1}}{s_n} & \cdots & \dfrac{\sigma\gamma_{np}}{s_n} & \dfrac{\varepsilon\gamma_{n,p+1}}{s_n} & \cdots & \dfrac{\varepsilon\gamma_{nn}}{s_n}
\end{array}\right). \tag{2.10}
$$

因为相似变换不改变矩阵的特征值，所以可以转为考察(2.10)所示的矩阵的特征值分布. 按照 Gerschgorin 定理，$D^{-1}Z^H(A+\varepsilon B)XD$ 的 Gerschgorin 圆盘如下:

当 $i=1,\cdots,p$ 时

$$G_i=\left\{z\in\mathbf{C}:\left|z-\left(\lambda_1+\frac{\varepsilon\gamma_{ii}}{s_i}\right)\right|\right.$$

$$\left.\leqslant\varepsilon\sum_{\substack{j=1\\j\neq i}}^{p}\frac{|\gamma_{ij}|}{s_i}+\varepsilon^2\sum_{i=p+1}^{n}\frac{|\gamma_{ij}|}{\sigma s_i}\right\};\tag{2.11}$$

当 $i=p+1,\cdots,n$ 时

$$G_i=\left\{z\in\mathbf{C}:\left|z-\left(\lambda_i+\frac{\varepsilon\gamma_{ii}}{s_i}\right)\right|\right.$$

$$\left.\leqslant\sum_{j=1}^{p}\frac{|\sigma\gamma_{ij}|}{s_i}+\varepsilon\sum_{\substack{i=p+1\\j\neq i}}^{n}\frac{|\gamma_{ij}|}{s_i}\right\}.$$

注意到，当 $i=1,\cdots,p$ 时

$$G_i\subseteq\left\{z\in\mathbf{C}:|z-\lambda_1|\leqslant\frac{np\varepsilon+n(n-p)\varepsilon^2/\sigma}{s_i}\right\},$$

而当 $i=p+1,\cdots,n$ 时

$$G_i\subseteq\left\{z\in\mathbf{C}:|z-\lambda_i|\leqslant\frac{np\sigma+n(n-p)\varepsilon}{s_i}\right\};$$

所以只要取 σ，使得对于 $1\leqslant i\leqslant p$ 和 $p+1\leqslant j\leqslant n$，恒有

$$\frac{np\varepsilon+n(n-p)\varepsilon^2/\sigma}{s_i}+\frac{np\sigma+n(n-p)\varepsilon}{s_j}$$

$$<\min_{p+1\leqslant k\leqslant n}|\lambda_1-\lambda_k|\equiv\delta,\tag{2.12}$$

则 $\bigcup_{i=1}^{p}G_i$ 与其它的 Gerschgorin 圆盘分离．比如，可取

$$\sigma=\min_{p+1\leqslant j\leqslant n}s_j\cdot\frac{\delta}{2np},$$

则当 ε 充分小时，(2.12)必成立．这时在 $\bigcup_{i=1}^{p}G_i$ 内恰有 $A+\varepsilon B$ 的 p 个特征值 $\mu_1(\varepsilon),\cdots,\mu_p(\varepsilon)$．由(2.11)知，这些 $\mu_i(\varepsilon)$ $(1\leqslant i\leqslant p)$ 与 λ_1 的距离为

$$|\mu_i(\varepsilon)-\lambda_1| \leqslant \left(np\varepsilon+\frac{n(n-p)\varepsilon^2}{\sigma}\right)\Big/\min_{1\leqslant i\leqslant p} s_i$$

$$\approx \frac{np\varepsilon}{\min\limits_{1\leqslant i\leqslant p} s_i},\quad i=1,\cdots,p. \tag{2.13}$$

特别地，当 $p=1$ 时，只要 ε 足够小，则 $A+\varepsilon B$ 必有一特征值 $\mu_1(\varepsilon)$，满足

$$|\mu_1(\varepsilon)-\lambda_1| \leqslant \left(n\varepsilon+\frac{n(n-1)\varepsilon^2}{\sigma}\right)\Big/ s_1 \approx \frac{n\varepsilon}{s_1}. \tag{2.14}$$

Wilkinson 曾利用 Gerschgorin 定理，对不同特征值结构的情形，进行了细致的扰动分析，其步骤为：①把原来的矩阵化成 Jordan 标准形，②用对角相似变换，使扰动后的矩阵包含所要考虑的特征值的 Gerschgorin 圆盘半径减小，③利用 Gerschgorin 定理作出估计。

在特征值的扰动分析中，利用 Gerschgorin 定理往往是有效的。

习题

1. 设

$$A=\begin{pmatrix} 1 & \alpha_{12} & \cdots & & \alpha_{1n} \\ \alpha_{21} & 0 & \alpha_{23} & \cdots & \alpha_{2n} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \alpha_{n-1,n} \\ \vdots & & & \ddots & \ddots \\ \alpha_{n1} & \cdots & \alpha_{n,n-1} & & 0 \end{pmatrix},\quad |\alpha_{ij}|\leqslant\varepsilon.$$

其中 ε 满足 $[(n-2)+2\sqrt{n-1}]\varepsilon<1$。 试证：$A$ 恰有一个特征值落入圆盘

$$\mathscr{D}_1=\{z\in\mathbf{C}: |z-1|\leqslant\sqrt{n-1}\,\varepsilon\}.$$

2. 设

$$A=\begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}\in\mathbf{C}^{m\times m},$$

$$B=(\beta_{ij})\in\mathbb{C}^{m\times m},\ |\beta_{ij}|\leqslant 1,\ 1\leqslant i,j\leqslant m.$$

令

$$C_\varepsilon=A+\varepsilon B,\ 0<\varepsilon<1.$$

试证：存在常数 K，使得

$$|\lambda_i(\varepsilon)-\lambda|\leqslant K\varepsilon^{\frac{1}{m}},\ \forall\lambda_i(\varepsilon)\in\lambda(C_\varepsilon).$$

3. 设 $A=(\alpha_{ij})\in\mathbb{C}^{n\times n}$，满足

$$|\alpha_{ii}|>\sum_{j\neq i}|\alpha_{ij}|\equiv\rho_i, i=1,\cdots,n.$$

试证

$$|\det A|\geqslant\prod_{i=1}^{n}(|\alpha_{ii}|-\rho_i).$$

§3 Hermite 阵的特征值

3.1 极小极大定理

首先证明一条引理.

引理 3.1. 设 $\mathscr{X}$ 与 $\mathscr{Y}$ 是 $\mathbb{C}^n$ 内的子空间. 如果 $\dim(\mathscr{X})>\dim(\mathscr{Y})$，则必存在非零向量 $\boldsymbol{x}\in\mathscr{X}$，使得 $x\perp\mathscr{Y}$.

证明:

分别在 $\mathscr{X}$ 与 $\mathscr{Y}$ 内取基底向量 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_l$ 与 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_m$.

令

$$X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_l), Y=(\boldsymbol{y}_1,\cdots,\boldsymbol{y}_m), l>m.$$

于是 $\mathscr{X}$ 中任一向量 $\boldsymbol{x}$ 可以表示成 $\boldsymbol{x}=X\boldsymbol{u}$，$\boldsymbol{u}\in\mathbb{C}^l$. 为使 $\boldsymbol{x}\perp\mathscr{Y}$，必须而且只需

$$Y^HX\boldsymbol{u}=0, \tag{3.1}$$

其中 $Y^HX\in\mathbb{C}^{m\times l}$. 注意到方程组(3.1)中方程的个数 m 小于未知数个数 l，所以 (3.1) 必有非零解 $\boldsymbol{u}$. 取之构造 $\boldsymbol{x}=X\boldsymbol{u}$，即满足 $\boldsymbol{x}\perp\mathscr{Y}$. □

定理 3.1 (极小极大定理). 设 $A\in\mathbb{C}^{n\times n}$ 为 Hermite 阵，其

特征值为 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. 则有

$$
\left.\begin{aligned}
\lambda_i &= \max_{\substack{\mathscr{X} \\ \dim(\mathscr{X})=i}} \min_{\substack{x \in \mathscr{X} \\ x \neq 0}} \frac{\boldsymbol{x}^H A \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{x}} \\
&= \max_{\substack{\mathscr{X} \\ \dim(\mathscr{X})=i}} \min_{\substack{\boldsymbol{u} \in \mathscr{X} \\ \|\boldsymbol{u}\|_2=1}} \boldsymbol{u}^H A \boldsymbol{u}
\end{aligned}\right\} \tag{3.2}
$$

和

$$
\left.\begin{aligned}
\lambda_i &= \min_{\substack{\mathscr{X} \\ \dim(\mathscr{X})=n-i+1}} \max_{\substack{x \in \mathscr{X} \\ x \neq 0}} \frac{\boldsymbol{x}^H A \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{x}} \\
&= \min_{\substack{\mathscr{X} \\ \dim(\mathscr{X})=n-i+1}} \max_{\substack{\boldsymbol{u} \in \mathscr{X} \\ \|\boldsymbol{u}\|_2=1}} \boldsymbol{u}^H A \boldsymbol{u}.
\end{aligned}\right\} \tag{3.3}
$$

其中 $i = 1, \cdots, n$.

证明:

因为通过考虑 $-A$, 即可由(3.2)导出 (3.3), 以下证明(3.2). 取 A 的酉-对角分解:

$$A = U\Lambda U^H,$$

其中 $U = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)$ 为酉阵, $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, $\lambda_1 \geqslant \cdots \geqslant \lambda_n$.

考虑 $\mathbf{C}^n$ 的任一 i 维子空间 $\mathscr{X}$. 据引理 3.1, 存在非零向量 $\boldsymbol{x} \in \mathscr{X}$, 使得 $\boldsymbol{x} \perp R(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{i-1})$, $R(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{i-1})$ 表示由 $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{i-1}$ 张成的子空间,因而

$$U^H \boldsymbol{x} = (0, \cdots, 0, \xi_i, \cdots, \xi_n)^T;$$

于是有

$$
\begin{aligned}
\frac{\boldsymbol{x}^H A \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{x}} &= \frac{(U^H \boldsymbol{x})^H \Lambda (U^H \boldsymbol{x})}{(U^H \boldsymbol{x})^H (U^H \boldsymbol{x})} = \frac{\lambda_i |\xi_i|^2 + \cdots + \lambda_n |\xi_n|^2}{|\xi_i|^2 + \cdots + |\xi_n|^2} \\
&\leqslant \lambda_i
\end{aligned}
$$

和

$$\min_{\substack{\boldsymbol{x} \in \mathscr{X} \\ \boldsymbol{x} \neq 0}} \frac{\boldsymbol{x}^H A \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{x}} \leqslant \lambda_i. \tag{3.4}$$

另一方面,取 $\mathscr{X}' = \mathscr{R}(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_i)$, $\mathscr{X}'$ 中的 $\boldsymbol{x}$ 必可表示成 $\boldsymbol{x} = \xi_1 \boldsymbol{u}_1 + \cdots + \xi_i \boldsymbol{u}_i$, 因而

$$\min_{\substack{x \in \mathscr{X}' \\ x \neq 0}} \frac{x^H A x}{x^H x} = \min_{\substack{(\xi_1,\cdots,\xi_i)^T \in \mathbb{C}^i \\ (\xi_1,\cdots,\xi_i)^T \neq 0}} \frac{\lambda_1|\xi_1|^2 + \cdots + \lambda_i|\xi_i|^2}{|\xi_1|^2 + \cdots + |\xi_i|^2}$$

$$= \lambda_i. \tag{3.5}$$

把(3.4)与(3.5)联系起来，便得出(3.2). □

从 Hermite 阵的极小极大定理可以导出一些重要的结论.

分别在(3.2)和(3.3)中取 i 为 1 和 n，则得到

推论 3.1. 设 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 阵，它的最大与最小特征值分别为 λ_1 与 λ_n. 则

$$\lambda_1 = \max_{\substack{u \in \mathbb{C}^n \\ \|u\|_2 = 1}} u^H A u, \quad \lambda_n = \min_{\substack{u \in \mathbb{C}^n \\ \|u\|_2 = 1}} u^H A u.$$

定理 3.2 (分隔定理). 设 $A \in \mathbb{C}^{n\times n}$ 为 Hermite 阵，$U_{n-1} \in \mathbb{C}^{n\times(n-1)}$，$U_{n-1}^H U_{n-1} = I^{(n-1)}$. 令

$$A' = U_{n-1}^H A U_{n-1}.$$

A 与 A' 的特征值分别为 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$ 与 $\lambda_1(A') \geqslant \cdots \geqslant \lambda_{n-1}(A')$. 则

$$\lambda_1(A) \geqslant \lambda_1(A') \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_{n-1}(A) \geqslant \lambda_{n-1}(A') \geqslant \lambda_n(A). \tag{3.6}$$

证明:

据定理 3.1，有

$$\lambda_i(A') = \max_{\substack{\mathscr{Y} \\ \dim(\mathscr{Y}) = i}} \min_{\substack{v \in \mathscr{Y} \\ \|v\|_2 = 1}} v^H A' v \tag{3.7}$$

$$= \min_{\substack{\mathscr{Y} \\ \dim(\mathscr{Y}) = (n-1)-i+1}} \max_{\substack{v \in \mathscr{Y} \\ \|v\|_2 = 1}} v^H A' v. \tag{3.8}$$

设 $\hat{\mathscr{Y}} \subset \mathbb{C}^{n-1}$ 是等式(3.7)中的极大化子空间，于是有

$$\lambda_i(A') = \min_{\substack{v \in \hat{\mathscr{Y}} \\ \|v\|_2 = 1}} v^H A' v = \min_{\substack{v \in \hat{\mathscr{Y}} \\ \|v\|_2 = 1}} (U_{n-1}v)^H A (U_{n-1}v)$$

$$= \min_{\substack{u \in U_{n-1}\hat{\mathscr{Y}} \\ \|u\|_2 = 1}} u^H A u \leqslant \max_{\substack{\mathscr{X} \\ \dim(\mathscr{X}) = i}} \min_{\substack{u \in \mathscr{X} \\ \|u\|_2 = 1}} u^H A u$$

$$= \lambda_i(A), \quad i = 1, \cdots, n-1; \tag{3.9}$$

另一方面，设 $\check{\mathscr{Y}}$ 是等式(3.8)中的极小化子空间，于是有

$$\lambda_i(A') = \max_{\substack{v\in\widetilde{\mathscr{Y}}\\ \|v\|_2=1}} v^H A' v = \max_{\substack{v\in\widetilde{\mathscr{Y}}\\ \|v\|_2=1}} (U_{n-1}v)^H A (U_{n-1}v)$$

$$= \max_{\substack{u\in U_{n-1}\widetilde{\mathscr{Y}}\\ \|u\|_2=1}} u^H A u \geqslant \min_{\substack{\mathscr{X}\\ \dim(\mathscr{X})=n-(i+1)+1}} \max_{\substack{u\in\mathscr{X}\\ \|u\|_2=1}} u^H A u$$

$$= \lambda_{i+1}(A),\ i = 1,\cdots,n-1. \tag{3.10}$$

(3.9)与(3.10)给出了(3.6). □

下述定理是推论 3.1 的推广.

定理 3.3. 设 $A\in\mathbb{C}^{n\times n}$ 是 Hermite 阵,其特征值为 $\lambda_1\geqslant\cdots\geqslant\lambda_n$. 又设 $U_m\in\mathbb{C}^{n\times m}$, $U_m^H U_m = I^{(m)}$. 则有

$$\lambda_1+\cdots+\lambda_m = \max_{U_m} \operatorname{tr}(U_m^H A U_m) \tag{3.11}$$

和

$$\lambda_{n-m+1}+\cdots+\lambda_n = \min_{U_m} \operatorname{tr}(U_m^H A U_m). \tag{3.12}$$

证明:

在 U_m 右边补上 $n-m-1$ 列,构成 U_{n-1}, 满足 $U_{n-1}^H U_{n-1} = I^{(n-1)}$. 据定理 3.2, 有

$$\lambda_i(A) \geqslant \lambda_i(U_{n-1}^H A U_{n-1})$$

$$\geqslant \lambda_i\left(\left(U_{n-1}\begin{pmatrix} I^{(n-2)}\\ 0\end{pmatrix}\right)^H \cdot A\left(U_{n-1}\begin{pmatrix} I^{(n-2)}\\ 0\end{pmatrix}\right)\right)$$

$$\geqslant\cdots\geqslant \lambda_i(U_m^H A U_m),\ i=1,\cdots,m;$$

由此得到

$$\lambda_1+\cdots+\lambda_m \geqslant \operatorname{tr}(U_m^H A U_m). \tag{3.13}$$

同时,有

$$\lambda_i(A) \leqslant \lambda_{i-1}(U_{n-1}^H A U_{n-1})$$

$$\leqslant \lambda_{i-2}\left(\left(U_{n-1}\begin{pmatrix} I^{(n-2)}\\ 0\end{pmatrix}\right)^H A\left(U_{n-1}\begin{pmatrix} I^{(n-2)}\\ 0\end{pmatrix}\right)\right)$$

$$\leqslant\cdots\leqslant \lambda_{i-(n-m)}(U_m^H A U_m),$$

$$i = n, n-1,\cdots, n-m+1;$$

由此得到

$$\lambda_{n-m+1}+\cdots+\lambda_n \leqslant \operatorname{tr}(U_m^H A U_m). \tag{3.14}$$

另一方面,分别取 U_m 为 A 属于 $\lambda_1,\cdots,\lambda_m$ 和 $\lambda_{n-m+1},\cdots,\lambda_n$

的单位特征向量所构成的列正交阵，可知(3.13)和(3.14)中的等式能够成立．所以(3.11)与(3.12)成立．□

3.2 极小极大定理的一般形式

Wielandt[187] 把定理 3.1 推广成了相当一般的形式．

首先证明两个引理．

引理 3.2. 设 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 阵．则 $A=0$ 的必要充分条件是

$$\boldsymbol{x}^H A\boldsymbol{x}=0,\quad \forall \boldsymbol{x}\in \mathbb{C}^n. \tag{3.15}$$

证明:

显然只需证明条件(3.15)的充分性．

取 A 的酉-对角分解：$A=U^H\Lambda U$，其中 U 为酉阵，$\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$．则(3.15)便是

$$(U\boldsymbol{x})^H\Lambda(U\boldsymbol{x})=0,\ \forall \boldsymbol{x}\in \mathbb{C}^n$$

即

$$\boldsymbol{y}^H\Lambda \boldsymbol{y}=0,\ \forall \boldsymbol{y}\in \mathbb{C}^n.$$

由此立即导出 $\Lambda=0$，因而 $A=0$．□

引理 3.3. 设 $A\in\mathbb{C}^{n\times n}$ 是 Hermite 阵，$\lambda(A)=\{\lambda_i\}$ 满足 $\lambda_1\geqslant\cdots\geqslant\lambda_n$．如果在 $\mathbb{C}^n$ 的一个 $n-1$ 维子空间 $\mathscr{X}$ 上，按照

$$\mathscr{A}'\boldsymbol{x}=P_{\mathscr{X}}A\boldsymbol{x},\quad \forall \boldsymbol{x}\in\mathscr{X} \tag{3.16}$$

定义一个线性变换 $\mathscr{A}'$，则下述结论成立：

1) 在 $\mathscr{X}$ 内任取一组标准正交基，如果 $\mathscr{A}'$ 在该组标准正交基之下的矩阵表示记作 A'，则 A' 必为 $n-1$ 行列的 Hermite 阵；

2) A' 的特征值 $\lambda_1'\geqslant\cdots\geqslant\lambda_{n-1}'$ 必满足

$$\lambda_1\geqslant\lambda_1'\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_{n-1}\geqslant\lambda_{n-1}'\geqslant\lambda_n. \tag{3.17}$$

证明:

1) 首先在 $\mathbb{C}^n$ 内取自然基底 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$，其中 $\boldsymbol{e}_i$ 表示 $I^{(n)}$ 的第 i 列向量，$i=1,\cdots,n$．再在 $\mathscr{X}$ 中任意选取一组标准正交基 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n-1}$，然后取单位向量 $\boldsymbol{v}_n$，使 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n$ 构成$\mathbb{C}^n$

的一组标准正交基，其中

$$\boldsymbol{v}_i = \sum_{j=1}^{n} v_{ji}\boldsymbol{e}_j = (v_{1i}, \cdots, v_{ni})^T, i = 1, \cdots, n.$$

记

$$V_{n-1} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-1}), V = (V_{n-1}, \boldsymbol{v}_n), \tag{3.18}$$

显然 $V \in \mathbb{C}^{n\times n}$ 是酉阵，V_{n-1} 满足

$$V_{n-1}^H V_{n-1} = I^{(n-1)}, V_{n-1} \in \mathbb{C}^{n\times(n-1)}. \tag{3.19}$$

把线性变换 $\mathscr{A}'$ 在基底 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-1}$ 下的矩阵表示记作 A'，即

$$\mathscr{A}'V_{n-1} = V_{n-1}A', \quad A' \in \mathbb{C}^{(n-1)\times(n-1)}. \tag{3.20}$$

根据(3.16)，对于 $\mathscr{X}$ 中的任一向量 $\boldsymbol{x} = V_{n-1}\boldsymbol{x}'$，$\boldsymbol{x}' \in \mathbb{C}^{n-1}$，有

$$(V_{n-1}\boldsymbol{x}')^H(\mathscr{A}'V_{n-1}\boldsymbol{x}') = (V_{n-1}\boldsymbol{x}')^H A(V_{n-1}\boldsymbol{x}'),$$
$$\forall \boldsymbol{x}' \in \mathbb{C}^{n-1}. \tag{3.21}$$

由(3.20)可知(3.21)式的左端为

$$(V_{n-1}\boldsymbol{x}')^H V_{n-1}A'\boldsymbol{x}' = \boldsymbol{x}'^H A'\boldsymbol{x}'; \tag{3.22}$$

代入(3.21)，得到

$$\boldsymbol{x}'^H A'\boldsymbol{x}' = \boldsymbol{x}'^H V_{n-1}^H A V_{n-1}\boldsymbol{x}', \ \forall \boldsymbol{x}' \in \mathbb{C}^{n-1}. \tag{3.23}$$

(3.23) 表示

$$\boldsymbol{x}'^H\left(\frac{A' - A'^H}{2i}\right)\boldsymbol{x}' = 0,$$

所以根据引理 3.2，A' 为 Hermite 阵；再由(3.23)可导出

$$A' = V_{n-1}^H A V_{n-1}. \tag{3.24}$$

2) 利用定理 3.2，从(3.19)和 (3.24) 立即得到不等式 (3.17)

□

定理 3.4 (Wielandt [187]). 设 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 阵，$\lambda(A) = \{\lambda_i\}$ 满足 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$. 又设 $i_1, \cdots, i_m$ 是适合

$$1 \leqslant i_1 < \cdots < i_m \leqslant n$$

的 m 个任意的自然数. 则有

$\lambda_{i_1} + \cdots + \lambda_{i_m}$

$$= \max_{\substack{\mathscr{X}_{i_1}\subset\cdots\subset\mathscr{X}_{i_m}\\ \dim(\mathscr{X}_{i_k})=i_k\\ (k=1,\cdots,m)}} \min_{\substack{U_m=(u_{i_1},\cdots,u_{i_m})\\ u_{i_k}\in\mathscr{X}_{i_k}(k=1,\cdots,m)\\ U_m^H U_m=I^{(m)}}} \operatorname{tr}(U_m^H A U_m) \tag{3.25}$$

和

$$\lambda_{i_1}+\cdots+\lambda_{i_m}$$
$$= \min_{\substack{\mathscr{X}_{n-i_m+1}\subset\cdots\subset\mathscr{X}_{n-i_1+1}\\ \dim(\mathscr{X}_{n-i_k+1})=n-i_k+1\\ (k=1,\cdots,m)}} \max_{\substack{U_m=(u_{n-i_m+1},\cdots,u_{n-i_1+1})\\ u_{n-i_k+1}\in\mathscr{X}_{n-i_k+1}(k=1,\cdots,m)\\ U_m^H U_m=I^{(m)}}} \operatorname{tr}(U_m^H A U_m). \tag{3.26}$$

证明:

首先指出,关系式(3.26)可由(3.25)导出. 事实上, 可考虑 $-A$. 记 $-\lambda_i=\mu_{n-i+1}$, $i=n,n-1,\cdots,1$, 则有 $\lambda(-A)=\{\mu_{n-i+1}\}_{i=1}^n$ 满足 $\mu_1\geqslant\cdots\geqslant\mu_n$. 于是对于在 $1,\cdots,n$ 中任意选取的 m 个自然数

$$n-i_m+1\leqslant\cdots\leqslant n-i_1+1,$$

应用公式(3.25),得到

$$\mu_{n-i_m+1}+\cdots+\mu_{n-i_1+1}$$
$$= \max_{\substack{\mathscr{X}_{n-i_m+1}\subset\cdots\subset\mathscr{X}_{n-i_1+1}\\ \dim(\mathscr{X}_{n-i_k+1})=n-i_k+1\\ (k=1,\cdots,m)}} \min_{\substack{U_m=(u_{n-i_m+1},\cdots,u_{n-i_1+1})\\ u_{n-i_k+1}\in\mathscr{X}_{n-i_k+1}(k=1,\cdots,m)\\ U_m^H U_m=I^{(m)}}} \operatorname{tr}(U_m^H A U_m).$$

将 $\mu_{n-i_k+1}=-\lambda_{i_k}(k=1,\cdots,m)$ 代入上式左端, 即可导出(3.26)式.

因此,只需证明等式(3.25).

容易看出,证明(3.25)式等价于证明下面的 A) 与 B):

A) 设 $\mathscr{X}_{i_1}\subset\cdots\subset\mathscr{X}_{i_m}$ 是 $\mathbf{C}^n$ 中任意一列子空间, $\dim(\mathscr{X}_{i_k})=i_k$, $k=1,\cdots,m$. 则存在向量 $u_{i_1},\cdots,u_{i_m}$, 其中每个 $u_{i_k}\in\mathscr{X}_{i_k}$, $k=1,\cdots,m$, 并且 $U_m=(u_{i_1},\cdots,u_{i_m})$ 满足 $U_m^H U_m=I^{(m)}$, 使得

$$\lambda_{i_1}+\cdots+\lambda_{i_m}\geqslant\operatorname{tr}(U_m^H A U_m). \tag{3.27}$$

B) 在 $\mathbf{C}^n$ 中存在一个子空间序列 $\mathscr{X}_{i_1}\subset\cdots\subset\mathscr{X}_{i_m}$,

$\dim(\mathscr{X}_{i_k})=i_k, k=1,\cdots,m$，使得对于任意一组向量 $\boldsymbol{u}_{i_1},\cdots,\boldsymbol{u}_{i_m}$，只要 $\boldsymbol{u}_{i_k}\in\mathscr{X}_{i_k}, k=1,\cdots,m$，并且 $U_m=(\boldsymbol{u}_{i_1},\cdots,\boldsymbol{u}_{i_m})$ 满足 $U_m^H U_m=I^{(m)}$，则必有

$$\lambda_{i_1}+\cdots+\lambda_{i_m}\leqslant \operatorname{tr}(U_m^H A U_m). \tag{3.28}$$

先证明 B)。

设 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$ 是 A 分别属于 $\lambda_1,\cdots,\lambda_n$ 的特征向量所构成的一组标准正交基。定义

$$\mathscr{X}_j=R(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_j),\quad j=i_1,\cdots,i_m.$$

$R(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_j)$ 表示由 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_j$ 张成的子空间。显然有 $\mathscr{X}_{i_1}\subset\cdots\subset\mathscr{X}_{i_m}$，$\dim(\mathscr{X}_{i_k})=i_k, k=1,\cdots,m$。于是，对于 $\mathscr{X}_{i_k}$ 中任一单位向量 $\boldsymbol{u}_{i_k}=\sum_{j=1}^{i_k}\mu_{jk}\boldsymbol{v}_j$，有

$$\begin{aligned}\boldsymbol{u}_{i_k}^H A\boldsymbol{u}_{i_k}&=\left(\sum_{j=1}^{i_k}\mu_{jk}\boldsymbol{v}_j\right)^H A\left(\sum_{j=1}^{i_k}\mu_{jk}\boldsymbol{v}_j\right)\\&=\sum_{j=1}^{i_k}|\mu_{jk}|^2\lambda_j\geqslant\left(\sum_{j=1}^{i_k}|\mu_{jk}|^2\right)\lambda_{i_k}=\lambda_{i_k},\quad k=1,\cdots,m.\end{aligned}$$

因此，若在 $\mathscr{X}_{i_k}$ 中任取一单位向量 $\boldsymbol{u}_{i_k}$，$k=1,\cdots,m$，并令 $U_m=(\boldsymbol{u}_{i_1},\cdots,\boldsymbol{u}_{i_m})$，则有

$$\lambda_{i_1}+\cdots+\lambda_{i_m}\leqslant\operatorname{tr}(U_m^H A U_m).$$

即不等式(3.28)成立。

下面证明 A)。

对于矩阵的阶数 n 应用数学归纳法。

首先注意到，当 $m=n$ 时，容易选出一组标准正交向量 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$，其中 $\boldsymbol{u}_k\in\mathscr{X}_k$，$k=1,\cdots,n$。令 $U_n=(\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n)$，显然有 $U_n^H U_n=I^{(n)}$，并且

$$\operatorname{tr}(U^H AU)=\operatorname{tr}(A)=\lambda_1+\cdots+\lambda_n.$$

即不等式(3.27)成立。特别地，当 $n=1$ 时，(3.27)式成立。

假定当矩阵的阶数为 $n-1$ 时，结论 A) 已成立；现考虑 n 阶矩阵的情形如下。

设

$$\mathscr{X}_{i_1}\subset\cdots\subset\mathscr{X}_{i_m}\quad(m<n)$$

是 $\mathbf{C}^n$ 中的任一子空间序列,分两种情况讨论:

A-I) 第一种情况: $i_m<n$.

这时显然可以找到一个 $n-1$ 维子空间 $\mathscr{X}\subset\mathbf{C}^n$, 使得 $\mathscr{X}_{i_m}\subseteq\mathscr{X}$. 然后,按照引理 3.3 中的(3.16)式,在 $\mathscr{X}$ 上定义一个线性变换 $\mathscr{A}'$. 在 $\mathscr{X}$ 中取一组标准正交基 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n-1}$, 使得

$$R(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{i_k})=\mathscr{X}_{i_k},\ k=1,\cdots,m.$$

把 $\mathscr{A}'$ 在基底 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n-1}$ 之下的矩阵表示记作 A', $\lambda(A')=\{\lambda'_i\}_{i=1}^{n-1}$, 并记 $V_{n-1}=(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{n-1})$. 根据引理 3.3, 对于 $\mathscr{X}$ 中任一单位向量 $\boldsymbol{x}=V_{n-1}\boldsymbol{x}'$, $\boldsymbol{x}'\in\mathbf{C}^{n-1}$, 有

$$\boldsymbol{x}^H(\mathscr{A}'\boldsymbol{x})=\boldsymbol{x}'^HA'\boldsymbol{x}'.\tag{3.29}$$

注意到,对于 $k=1,\cdots,m$, $\boldsymbol{x}=V_{n-1}\boldsymbol{x}'$ 建立了 $\mathbf{C}^n$ 的子空间 $\mathscr{X}_{i_k}$ 与 $\mathbf{C}^{n-1}$ 的子空间

$$\mathscr{X}'_{i_k}=\{\boldsymbol{x}'=(\xi_1,\cdots,\xi_{i_k},0,\cdots,0)^T\in\mathbf{C}^{n-1}\}$$

的对应关系. $\mathbf{C}^{n-1}$ 中的子空间序列 $\{\mathscr{X}'_{i_k}\}_{k=1}^m$ 显然满足

$$\mathscr{X}'_{i_1}\subset\cdots\subset\mathscr{X}'_{i_m}\subseteq\mathbf{C}^{n-1},\ \dim(\mathscr{X}'_{i_k})=i_k,$$
$$k=1,\cdots,m.\tag{3.30}$$

根据归纳法假设, 存在向量 $\boldsymbol{u}'_{i_1},\cdots,\boldsymbol{u}'_{i_m}$, 其中每个 $\boldsymbol{u}'_{i_k}\in\mathscr{X}'_{i_k}$, $k=1,\cdots,m$, 并且 $U'_m=(\boldsymbol{u}'_{i_1},\cdots,\boldsymbol{u}'_{i_m})$ 满足 $U'^H_mU'_m=I^{(m)}$, 使得

$$\lambda'_{i_1}+\cdots+\lambda'_{i_m}\geqslant\mathrm{tr}(U'^H_mA'U'_m).\tag{3.31}$$

令

$$\boldsymbol{u}_{i_k}=V_{n-1}\boldsymbol{u}'_{i_k},k=1,\cdots,m$$

和

$$U_m=(\boldsymbol{u}_{i_1},\cdots,\boldsymbol{u}_{i_m})=V_{n-1}U'_m.\tag{3.32}$$

显然 $\boldsymbol{u}_{i_k}\in\mathscr{X}_{i_k},k=1,\cdots,m$, 并且 $U^H_mU_m=I^{(m)}$. 根据(3.32)和引理 3.3 中的(3.24),有

$$\mathrm{tr}(U^H_mAU_m)=\mathrm{tr}(U'^H_mV^H_{n-1}AV_{n-1}U'_m)=\mathrm{tr}(U'^H_mA'U'_m),$$

代入(3.31)式右端，得到

$$\lambda'_{i_1} + \cdots + \lambda'_{i_m} \geqslant \operatorname{tr}(U_m^H A U_m). \tag{3.33}$$

再根据引理 3.3，

$$\lambda_{i_k} \geqslant \lambda'_{i_k}, \quad k = 1, \cdots, m.$$

代入(3.33)式左端，便导出了不等式(3.27)。

A-II) 第二种情况: $i_m = n (m < n)$.

假设 $i_m = n$, $i_{m-1} = n-1, \cdots, i_{m-p} = n-p (p \geqslant 0)$, 而 $n-p-1 \notin \{i_1, \cdots, i_m\}$. 因此，若记 $m-p-1 = m'$，则

$$i_{m'} = i_{m-p-1} \leqslant n-p-2.$$

考虑 $\mathbf{C}^n$ 中的任一子空间序列

$$\mathscr{X}_{i_1} \subset \cdots \subset \mathscr{X}_{i_{m'}} \subset \mathscr{X}_{n-p} \subset \cdots \subset \mathscr{X}_n, \ m' = m-p-1,$$

其中

$$\dim(\mathscr{X}_k) = k, \ k = i_1, \cdots, i_{m'}, n-p, \cdots, n.$$

设 $\boldsymbol{w}_{n-p}, \boldsymbol{w}_{n-p+1}, \cdots, \boldsymbol{w}_n$ 是 A 分别属于 $\lambda_{n-p}, \lambda_{n-p+1}, \cdots, \lambda_n$ 的单位特征向量. 因为 $\boldsymbol{w}_{n-p}, \boldsymbol{w}_{n-p+1}, \cdots, \boldsymbol{w}_n$ 与 $\mathscr{X}_{i_{m'}}$ 张成$\mathbf{C}^n$ 中的一个维数 $\leqslant i_{m'} + (p+1) \leqslant n-p-2+(p+1) = n-1$ 的子空间，所以存在一个子空间 $\widetilde{\mathscr{X}} \subset \mathbf{C}^n$，其维数 $\dim(\widetilde{\mathscr{X}}) = n-1$，使得

$$\mathscr{X}_{i_{m'}} \subset \widetilde{\mathscr{X}}, \ \boldsymbol{w}_{n-p}, \cdots, \boldsymbol{w}_n \in \widetilde{\mathscr{X}}.$$

于是有

$$\mathscr{X}_{i_{m'}} \subseteq \mathscr{X}_{n-p} \cap \widetilde{\mathscr{X}} \subseteq \cdots \subseteq \mathscr{X}_{n-1} \cap \widetilde{\mathscr{X}} \subset \widetilde{\mathscr{X}}.$$

注意到

$$\begin{aligned}\dim(\mathscr{X}_k \cap \widetilde{\mathscr{X}}) &= \dim(\mathscr{X}_k) + \dim(\widetilde{\mathscr{X}}) - \dim(\mathscr{X}_k \cup \widetilde{\mathscr{X}}) \\ &\geqslant k + (n-1) - n = k-1, \\ &\quad k = n-p, \cdots, n-1.\end{aligned}$$

因此在 $\mathbf{C}^n$ 中存在子空间 $\widetilde{\mathscr{X}}_{n-p-1}, \widetilde{\mathscr{X}}_{n-p}, \cdots, \widetilde{\mathscr{X}}_{n-2}$，使得

$$\mathscr{X}_{i_{m'}} \subset \widetilde{\mathscr{X}}_{n-p-1} \subseteq \mathscr{X}_{n-p} \cap \widetilde{\mathscr{X}}, \cdots, \widetilde{\mathscr{X}}_{n-2} \subseteq \mathscr{X}_{n-1} \cap \widetilde{\mathscr{X}}$$

和

$$\widetilde{\mathscr{X}}_{n-p-1} \subset \widetilde{\mathscr{X}}_{n-p} \subset \cdots \subset \widetilde{\mathscr{X}}_{n-2},$$

其中 $\dim(\widetilde{\mathscr{X}}_k) = k, k = n-p-1, n-p, \cdots, n-2$. 从而有

$$\mathscr{X}_{i_1} \subset \cdots \subset \mathscr{X}_{i_{m'}} \subset \widetilde{\mathscr{X}}_{n-p-1} \subset \cdots \subset \widetilde{\mathscr{X}}_{n-2} \subset \widetilde{\mathscr{X}},$$

$$\dim(\widetilde{\mathscr{X}}) = n - 1.$$

仿 A-I),按照引理 3.3 中的 (3.16)式,在 $\mathbf{C}^n$ 中的 $n-1$ 维子空间 $\widetilde{\mathscr{X}}$ 上,定义算子 $\mathscr{A}'$. 在 $\widetilde{\mathscr{X}}$ 中取一组标准正交基 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-1}$, 使到

$$\mathscr{R}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{i_k}) = \mathscr{X}_{i_k}, \ k = 1, \cdots, m',$$

$$\mathscr{R}(\mathscr{X}_{i_{m'}}, \boldsymbol{v}_{n-p-1}, \cdots, \boldsymbol{v}_{n-l}) = \widetilde{\mathscr{X}}_{n-l}, \ l = p+1, \cdots, 2$$

和

$$\mathscr{R}(\widetilde{\mathscr{X}}_{n-2}, \boldsymbol{v}_{n-1}) = \widetilde{\mathscr{X}}.$$

记 $\mathscr{A}'$ 在 $\widetilde{\mathscr{X}}$ 的标准正交基 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-1}$ 之下的矩阵表示为 A', $\lambda(A') = \{\lambda'_i\}_{i=1}^{n-1}$, 令

$$V_{n-1} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-1}).$$

则仿 A-I)同样可证(根据归纳法假设): 存在向量 $\boldsymbol{u}'_{i_1}, \cdots, \boldsymbol{u}'_{i_{m'}}$, $\boldsymbol{u}'_{n-p-1}, \cdots, \boldsymbol{u}'_{n-1}$, 其中 $\boldsymbol{u}'_{i_k} \in \mathscr{X}'_{i_k}$, $k = 1, \cdots, m'$, $\boldsymbol{u}'_{n-l} \in \widetilde{\mathscr{X}}_{n-l}$, $l = p+1, \cdots, 2$, $\boldsymbol{u}'_{n-1} \in \widetilde{\mathscr{X}}$, 并且

$$U'_m = (\boldsymbol{u}'_{i_1}, \cdots, \boldsymbol{u}'_{i_{m'}}, \boldsymbol{u}'_{n-p-1}, \cdots, \boldsymbol{u}'_{n-1})$$

满足 $U'^H_m U'_m = I^{(m)}$, 使得

$$\lambda'_{i_1} + \cdots + \lambda'_{i_{m'}} + \lambda'_{n-p-1} + \cdots + \lambda'_{n-1} \geqslant \operatorname{tr}(U'^H_m A' U'_m). \tag{3.34}$$

再令

$$\boldsymbol{u}_{i_k} = V_{n-1}\boldsymbol{u}'_{i_k}, \ k = 1, \cdots, m';$$

$$\boldsymbol{u}_{n-l} = V_{n-1}\boldsymbol{u}'_{n-l}, \ l = p+1, \cdots, 1$$

和

$$U_m = (\boldsymbol{u}_{i_1}, \cdots, \boldsymbol{u}_{i_{m'}}, \boldsymbol{u}_{n-p-1}, \cdots, \boldsymbol{u}_{n-1}),$$

显然有

$$\boldsymbol{u}_{i_k} \in \mathscr{X}_{i_k}, \ k = 1, \cdots, m';$$

$$\boldsymbol{u}_{n-l} \in \widetilde{\mathscr{X}}_{n-l}, l = p+1, \cdots, 2, \boldsymbol{u}_{n-1} \in \widetilde{\mathscr{X}}$$

和

$$U^H_m U_m = I^{(m)}.$$

同 A-I) 可证

$$\mathrm{tr}(U_m'^H A' U_m') = \mathrm{tr}(U_m^H A U_m). \tag{3.35}$$

根据引理 3,3，有

$$\lambda_{i_k} \geqslant \lambda'_{i_k} \quad k = 1, \cdots, m'. \tag{3.36}$$

将(3.35)与(3.36)代入不等式(3.34)，便得到

$$\lambda_{i_1} + \cdots + \lambda_{i_{m'}} + \lambda'_{n-p-1} + \cdots + \lambda'_{n-1} \geqslant \mathrm{tr}(U_m^H A U_m). \tag{3.37}$$

因此，为了证明不等式(3.27)成立，只需再证明

$$\lambda_{n-p} + \cdots + \lambda_n \geqslant \lambda'_{n-p-1} + \cdots + \lambda'_{n-1}. \tag{3.38}$$

据已设，向量 $\boldsymbol{w}_{n-p}, \cdots, \boldsymbol{w}_n \in \tilde{\mathscr{R}}$ 适合

$$A\boldsymbol{w}_{n-l} = \lambda_{n-l}\boldsymbol{w}_{n-l}, \|\boldsymbol{w}_{n-l}\|_2 = 1, l = p, p-1, \cdots, 0. \tag{3.39}$$

显然对于 $\tilde{\mathscr{R}}$ 的标准正交基 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-1}, \boldsymbol{w}_{n-l}$ 可以表示成

$$\boldsymbol{w}_{n-l} = V_{n-1}\boldsymbol{w}'_{n-l}, \ \boldsymbol{w}'_{n-l} \in \mathbf{C}^{n-1}, l = p, p-1, \cdots, 0.$$

因而，代入 (3.39)，有

$$AV_{n-1}\boldsymbol{w}'_{n-l} = \lambda_{n-l}V_{n-1}\boldsymbol{w}'_{n-l}$$

即

$$A'\boldsymbol{w}'_{n-l} = \lambda_{n-l}\boldsymbol{w}'_{n-l}, \ l = p, \ p-1, \cdots, 0.$$

这表示 $\lambda_{n-p}, \lambda_{n-p+1}, \cdots, \lambda_n$ 也是 A' 的特征值；而且根据定理 3.2，它们只可能是 A' 的最小的 $p+1$ 个特征值．因此

$$\lambda_{n-p} + \lambda_{n-p+1} + \cdots + \lambda_n = \lambda'_{n-p-1} + \lambda'_{n-p} + \cdots + \lambda'_{n-1}.$$

即(3.38)成立．代入(3.37)，立即导出不等式(3.27)．□

注 3.1．在定理 3.4 中，取 $m=1$，便得到定理 3.1；在(3.25)中取 $i_k = k, k = 1, \cdots, m$，在(3.26)中取 $i_k = n-k+1$，$k = 1, \cdots, m$，则分别得到定理 3.3 的公式(3.11)和(3.12)．

注 3.2．在 Wielandt 之后，Amir Moéz 对 Hermite 阵的极小极大定理又作了一些推广．对此有兴趣的读者，可参阅他 1956 年的论文〔39〕．

3.3 Hermite 扰动

定理 3.5． 设 $A, B = A+E \in \mathbf{C}^{n\times n}$，$A, B$ 和 E 皆为 Hermite 阵，它们的特征值分别为

$$\lambda_1 \geqslant \cdots \geqslant \lambda_n, \mu_1 \geqslant \cdots \geqslant \mu_n \tag{3.40}$$

和

$$\varepsilon_1 \geqslant \cdots \geqslant \varepsilon_n. \tag{3.41}$$

则

$$\lambda_i + \varepsilon_n \leqslant \mu_i \leqslant \lambda_i + \varepsilon_1, i = 1, \cdots, n. \tag{3.42}$$

证明:

设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 是由 A 的特征向量组成的标准正交基，对应的特征值分别为 $\lambda_1, \cdots, \lambda_n$. 令 $\mathscr{X} = R(\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \cdots, \boldsymbol{x}_n)$. 于是由(3.3)和推论 3.1，得到

$$\begin{aligned}\mu_i &\leqslant \max_{\substack{\boldsymbol{u} \in \mathscr{X} \\ \|\boldsymbol{u}\|_2 = 1}} \boldsymbol{u}^H B \boldsymbol{u} \leqslant \max_{\substack{\boldsymbol{u} \in \mathscr{X} \\ \|\boldsymbol{u}\|_2 = 1}} \boldsymbol{u}^H A \boldsymbol{u} + \max_{\substack{\boldsymbol{u} \in \mathscr{X} \\ \|\boldsymbol{u}\|_2 = 1}} \boldsymbol{u}^H E \boldsymbol{u} \\ &= \lambda_i + \max_{\substack{\boldsymbol{u} \in \mathscr{X} \\ \|\boldsymbol{u}\|_2 = 1}} \boldsymbol{u}^H E \boldsymbol{u} \leqslant \lambda_i + \varepsilon_1, i = 1, \cdots, n.\end{aligned} \tag{3.43}$$

另一方面，令 $D = -E$，记 $\lambda(D) = \{\delta_i\}$，$\delta_1 \geqslant \cdots \geqslant \delta_n$，则 $A = B + D$. 根据(3.43)，有

$$\lambda_i \leqslant \mu_i + \delta_1, \ i = 1, \cdots, n,$$

其中 $\delta_1 = -\varepsilon_n$. 因此(3.42)成立. □

由(3.42)可知

$$|\mu_i - \lambda_i| \leqslant \max\{\varepsilon_1, |\varepsilon_n|\}, i = 1, \cdots, n.$$

再考虑到

$$\|E\|_2 = \max\{\varepsilon_1, |\varepsilon_n|\},$$

所以有下述重要结论(见[185]和[198]):

定理 3.6 (Weyl-Лидский 定理). 设 A, $B \in \mathbb{C}^{n \times n}$ 皆为 Hermite 阵，它们的特征值如(3.40)所示. 则

$$|\mu_i - \lambda_i| \leqslant \|B - A\|_2, \quad i = 1, \cdots, n; \tag{3.44}$$

即

$$v(A, B) \leqslant \|B - A\|_2. \tag{3.45}$$

下述定理是定理 3.5 的推广.

定理 3.7. 设 $A, B = A + E \in \mathbb{C}^{n \times n}$ 皆为 Hermite 阵，A, B 和 E 的特征值分别如(3.40)和(3.41)所示. 则

$$\lambda_1 + \cdots + \lambda_m + \varepsilon_{n-m+1} + \cdots + \varepsilon_n \leqslant \mu_1 + \cdots + \mu_m$$
$$\leqslant \lambda_1 + \cdots + \lambda_m + \varepsilon_1 + \cdots + \varepsilon_m, \quad m = 1, \cdots, n. \quad (3.46)$$

定理 3.7 可以利用定理 3.3 加以证明;同时,也可作为下述结果的一个特例.

定理3.8. 设 $A, B = A+E \in \mathbf{C}^{n\times n}$, A, B 和 E 皆为 Hermite 阵,它们的特征值分别如(3.40)和(3.41)所示. 则对于在 $1, \cdots, n$ 中任意选取的 m 个自然数 $i_1, \cdots, i_m$, 有

$$\lambda_{i_1} + \cdots + \lambda_{i_m} + \varepsilon_{n-m+1} + \cdots + \varepsilon_n \leqslant \mu_{i_1} + \cdots + \mu_{i_m}$$
$$\leqslant \lambda_{i_1} + \cdots + \lambda_{i_m} + \varepsilon_1 + \cdots + \varepsilon_m. \quad (3.47)$$

证明:

无妨假设 $1 \leqslant i_1 < \cdots < i_m \leqslant n$. 根据定理 3.4, 在 $\mathbf{C}^n$ 中存在一个子空间序列 $\mathscr{X}_{i_1} \subset \cdots \subset \mathscr{X}_{i_m}$, 使得

$$\mu_{i_1} + \cdots + \mu_{i_m} = \min_{\substack{U_m = (u_{i_1}, \cdots, u_{i_m}) \\ u_{i_k} \in \mathscr{X}_{i_k} (k=1,\cdots,m) \\ U_m^H U_m = I^{(m)}}} \operatorname{tr}(U_m^H B U_m).$$

设向量 $v_{i_k} \in \mathscr{X}_{i_k}$, $k = 1, \cdots, m$, $V_m = (v_{i_1}, \cdots, v_{i_m})$ 满足 $V_m^H V_m = I^{(m)}$, 且使得

$$\operatorname{tr}(V_m^H A V_m) = \min_{\substack{U_m = (u_{i_1}, \cdots, u_{i_m}) \\ u_{i_k} \in \mathscr{X}_{i_k} (k=1,\cdots,m) \\ U_m^H U_m = I^{(m)}}} \operatorname{tr}(U_m^H A U_m).$$

由定理 3.4 知

$$\operatorname{tr}(V_m^H A V_m) \leqslant \lambda_{i_1} + \cdots + \lambda_{i_m};$$

此外,利用定理 3.3, 有

$$\operatorname{tr}(V_m^H E V_m) \leqslant \varepsilon_1 + \cdots + \varepsilon_m.$$

因此得到

$$\mu_{i_1} + \cdots + \mu_{i_m} \leqslant \operatorname{tr}(V_m^H B V_m) = \operatorname{tr}(V_m^H A V_m) + \operatorname{tr}(V_m^H E V_m)$$
$$\leqslant \lambda_{i_1} + \cdots \lambda_{i_m} + \varepsilon_1 + \cdots + \varepsilon_m.$$

即(3.47)的右端不等式成立 .

令 $D = -E$, 记 $\lambda(D) = \{\delta_i\}, \delta_1 \geqslant \cdots \geqslant \delta_n$, 有 $A = B + D$. 根据上述结果,有

$$\lambda_{i_1} + \cdots + \lambda_{i_m} \leqslant \mu_{i_1} + \cdots + \mu_{i_m} + \delta_1 + \cdots + \delta_m,$$

其中 $\delta_i = -\varepsilon_{n-i+1}, i = 1, \cdots, m$. 因此不等式(3.47)的左端不等式亦成立. □

注 3.3　在不等式(3.47)中取 $i_k = k, k = 1, \cdots, m$, 即得到不等式(3.46).

定理 3.9 (Hoffman-Wielandt 定理). 在定理 3.6 的假设下,有

$$e(A,B) = \sqrt{\sum_{i=1}^{n} (\mu_i - \lambda_i)^2} \leqslant \|B - A\|_F. \tag{3.48}$$

定理 3.9 是 § 4 中定理 4.3 的一个推论, 此处证略 (读者可参阅 Wilkinson [189]第二章的一个初等证明).

3.4　关于奇异值的扰动

定理 3.10[126]. 设 $A, B \in \mathbf{C}^{m\times n}$, 它们的奇异值分别是

$$\sigma_1 \geqslant \cdots \geqslant \sigma_n \geqslant 0, \tau_1 \geqslant \cdots \geqslant \tau_n \geqslant 0.$$

则对于 $\mathbf{C}^{m\times n}$ 上的任一酉不变范数$\|\ \ \|$,有

$$\|\operatorname{diag}(\tau_1 - \sigma_1, \cdots, \tau_n - \sigma_n)\| \leqslant \|B - A\|. \tag{3.49}$$

注 3.4　(3.49)式左端的范数是 $\mathbf{C}^{m\times n}$ 上的酉不变范数$\|\ \ \|$在 $\mathbf{C}^{n\times n}$ 上的限制, 即: 它们是借助于 $\mathbf{R}^n$ 上的同一个 SG 函数 Φ 得到的,因此仍记为$\|\ \ \|$.

定理 3.10 的证明:

因为当 $m > n$ 时,可分别用 $m \times m$ 方阵$(A, 0)$与$(B, 0)$代替 A 与 B,当 $m < n$ 时,可分别用 $n \times n$ 方阵$\begin{pmatrix} A \\ 0 \end{pmatrix}$与$\begin{pmatrix} B \\ 0 \end{pmatrix}$代替 A 与 B, 而这样做并不改变不等式(3.49)两端的酉不变范数的值; 所以,只需考虑 $m = n$ 的情形.

首先,将 $B - A = E$ 写成

$$\begin{pmatrix} 0 & B \\ B^H & 0 \end{pmatrix} - \begin{pmatrix} 0 & A \\ A^H & 0 \end{pmatrix} = \begin{pmatrix} 0 & E \\ E^H & 0 \end{pmatrix}. \tag{3.50}$$

设 E 的奇异值为 $\varepsilon_1 \geqslant \cdots \geqslant \varepsilon_n \geqslant 0$. 则(3.50)中的三个 Hermite

阵$\begin{pmatrix} 0 & B \\ B^H & 0 \end{pmatrix}$，$\begin{pmatrix} 0 & A \\ A^H & 0 \end{pmatrix}$和$\begin{pmatrix} 0 & E \\ E^H & 0 \end{pmatrix}$的特征值分别为

$$\left.\begin{array}{l} \tau_1 \geqslant \cdots \geqslant \tau_n \geqslant -\tau_n \geqslant \cdots \geqslant -\tau_1, \\ \sigma_1 \geqslant \cdots \geqslant \sigma_n \geqslant -\sigma_n \geqslant \cdots \geqslant -\sigma_1 \end{array}\right\} \tag{3.51}$$

和

$$\varepsilon_1 \geqslant \cdots \geqslant \varepsilon_n \geqslant -\varepsilon_n \geqslant \cdots \geqslant -\varepsilon_1. \tag{3.52}$$

(3.51) 和 (3.52)可如下看出．比如考虑矩阵 A．设 A 的奇异值分解为 $A = U\Sigma V^H$，其中 U 与 V 为酉阵，$\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$．直接验证可知

$$\begin{pmatrix} 0 & A \\ A^H & 0 \end{pmatrix} = \hat{U} \begin{pmatrix} \Sigma & 0 \\ 0 & -\Sigma \end{pmatrix} \hat{U}^H,$$

其中

$$\hat{U} = \begin{pmatrix} \frac{1}{\sqrt{2}} U & \frac{1}{\sqrt{2}} U \\ \frac{1}{\sqrt{2}} V & -\frac{1}{\sqrt{2}} V \end{pmatrix}$$

为 $2n$ 阶酉阵．

将(3.51)和(3.52)改写为

$$\tau_1 \geqslant \cdots \geqslant \tau_{2n}, \ \sigma_1 \geqslant \cdots \geqslant \sigma_{2n}, \ \varepsilon_1 \geqslant \cdots \geqslant \varepsilon_{2n},$$

则据(3.47)右端不等式，有

$$(\tau_{i_1} - \sigma_{i_1}) + \cdots + (\tau_{i_m} - \sigma_{i_m}) \leqslant \varepsilon_1 + \cdots + \varepsilon_m,$$

其中 $i_1, \cdots, i_m$ 是适合

$$1 \leqslant i_1 < \cdots < i_m \leqslant n$$

的 m 个任意的自然数，$m = 1, 2, \cdots, 2n$．由此可导出

$$|\tau_1 - \sigma_1| + \cdots + |\tau_i - \sigma_i| \leqslant \varepsilon_1 + \cdots + \varepsilon_i, i = 1, \cdots, n.$$

于是根据第二章 § 3 引理 3.10，对于 $\mathbb{R}^n$ 上任一SG 函数 Φ，必有

$$\Phi(\tau_1 - \sigma_1, \cdots, \tau_n - \sigma_n) \leqslant \Phi(\varepsilon_1, \cdots, \varepsilon_n). \tag{3.53}$$

设$\| \ \|$是 $\mathbb{C}^{n\times n}$ 上借助于 $\mathbb{R}^n$ 上某一SG 函数 Φ 得到的酉不变范数，则由 (3.53) 立即得到不等式(3.49)．□

下面两条重要结果是定理 3.10 的推论．

定理 3.11. 在定理 3.10 的假设下，有

$$|\tau_i - \sigma_i| \leqslant \|B - A\|_2, \quad i = 1, \cdots, n. \tag{3.54}$$

定理 3.12. 在定理 3.10 的假设下，有

$$\sqrt{\sum_{i=1}^{n} (\tau_i - \sigma_i)^2} \leqslant \|B - A\|_F. \tag{3.55}$$

定理 3.10—3.12 说明奇异值具有良好的稳定性质，这是矩阵的奇异值分解有着广泛的实际应用的理论根据之一。

习题

1. 设 $A \geqslant B \geqslant 0$，$\lambda(A) = \{\lambda_i\}_{i=1}^n$ 和 $\lambda(B) = \{\mu_i\}_{i=1}^n$ 满足 $\lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ 和 $\mu_1 \geqslant \cdots \geqslant \mu_n \geqslant 0$。试证 $\lambda_i \geqslant \mu_i$，$i = 1, \cdots, n$。

2. 设 $A \in \mathbb{C}^{m\times n}$ 有奇异值 $\sigma_1 \geqslant \cdots \geqslant \sigma_n \geqslant 0$。则

$$\sigma_i = \max_{\substack{\mathscr{X} \\ \dim(\mathscr{X})=i}} \min_{\substack{x\in\mathscr{X} \\ x\neq 0}} \frac{\|Ax\|_2}{\|x\|_2} = \min_{\substack{\mathscr{X} \\ \dim(\mathscr{X})=n-i+1}} \max_{\substack{x\in\mathscr{X} \\ x\neq 0}} \frac{\|Ax\|_2}{\|x\|_2};$$

特别地，有

$$\sigma_1 = \max_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2}, \quad \sigma_n = \min_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

此外，

$$\sigma_1 = \max_{\|x\|_2=\|y\|_2=1} |y^H A x|.$$

3. 设 $A = (\alpha_{ij}) \in \mathbb{C}^{n\times n}$ 是 Hermite 阵。证明：在每个圆盘

$$\mathscr{D}_i = \left\{ z \in \mathbb{C} : |z - \alpha_{ii}| \leqslant \sqrt{\sum_{j\neq i} |\alpha_{ij}|^2} \right\}$$

内，必有 A 的特征值，$i = 1, \cdots, n$。

4. 设 $A \in \mathbb{C}^{n\times n}$。记 $A = H_1 + iH_2$，其中

$$H_1 = \frac{1}{2}(A + A^H), H_2 = \frac{1}{2i}(A - A^H)$$

为 Hermite 阵. 设 $\lambda(H_1)=\{\alpha_i\}$, $\lambda(H_2)=\{\beta_i\}$, $\alpha_1\geqslant\cdots\geqslant\alpha_n$, $\beta_1\geqslant\cdots\geqslant\beta_n$. 试证

$$\alpha_1\geqslant \mathrm{Re}\lambda\geqslant\alpha_n,\ \beta_1\geqslant \mathrm{Im}\lambda\geqslant\beta_n,\ \forall\lambda\in\lambda(A).$$

5. 设

$$A=\begin{pmatrix}3 & 1.1 & -0.1\\ 0.9 & 15 & -2.1\\ 0.1 & -1.9 & 19.5\end{pmatrix}.$$

试利用上题和 Gerschgorin 定理,确定复平面上包含 A 的特征值的一个区域.

6. 设 A 为 Hermite 阵. 试证 A 为正定阵的必要与充分条件是 A 的顺序主子阵的行列式皆为正值. 此外,举例说明: 仅仅 A 的顺序主子阵的行列式皆为非负值,并不能保证 A 为半正定阵.

7. 设 $A\in\mathbb{C}^{n\times n}$ 为正定阵,其特征值为 $\lambda_1\geqslant\cdots\geqslant\lambda_n>0$. 任取 $X_1\in\mathbb{C}^{n\times m}(1\leqslant m\leqslant n)$. 试证

$$\lambda_1\cdots\lambda_m\det(X_1^HX_1)\geqslant\det(X_1^HAX_1)$$
$$\geqslant\lambda_{n-m+1}\cdots\lambda_n\det(X_1^HX_1).$$

8. 设 $A\in\mathbb{C}^{n\times n}$, $\lambda(A)=\{\lambda_i\}$, $\sigma(A)=\{\sigma_i\}$. 试证: 如果

$$|\lambda_1|\geqslant\cdots\geqslant|\lambda_n|,\ \sigma_1\geqslant\cdots\geqslant\sigma_n,$$

则

$$\prod_{i=1}^k\sigma_i\geqslant\prod_{i=1}^k|\lambda_i|,\ k=1,\cdots,n.$$

9. 设 $A,B\in\mathbb{C}^{n\times n}$, $C=AB$. A, B 与 C 的奇异值分别为

$$\alpha_1\geqslant\cdots\geqslant\alpha_n,\beta_1\geqslant\cdots\geqslant\beta_n,\gamma_1\geqslant\cdots\geqslant\gamma_n.$$

试证: 当 $i+j+1\leqslant n$ 时,有

$$\gamma_{i+j+1}\leqslant\alpha_{i+1}\beta_{j+1}.$$

(提示: ①分解 $C=WH$, W 为酉阵, H 为正定阵;

②注意到一个事实: 若 Hermite 阵 $A=U\Lambda U^H$, 其中 $U=(\boldsymbol{u}_1,\ \cdots,\ \boldsymbol{u}_n)$ 为酉阵, $\Lambda=\mathrm{diag}(\lambda_1,\ \cdots,\ \lambda_n)$, $\lambda_1\geqslant\cdots\geqslant\lambda_n$. 则

$$\lambda_{i+1} = \max_{\substack{x \perp \mathscr{R}(u_1,\cdots,u_i) \\ \|x\|_2=1}} x^H A x, \ i = 0,1,\cdots,n-1;$$

③利用 Cauchy 不等式

$$|x^H y|^2 \leqslant x^H x y^H y, \ \forall x, y \in \mathbf{C}^n;$$

④应用极小极大定理于 $(x^H H x)^2$.)

10. 设 $D = A + B \in \mathbf{C}^{n\times n}$. A, B 与D的奇异值分别为

$$\alpha_1 \geqslant \cdots \geqslant \alpha_n, \beta_1 \geqslant \cdots \geqslant \beta_n, \delta_1 \geqslant \cdots \geqslant \delta_n.$$

试证

$$\delta_{i+j+1} \leqslant \alpha_{i+1} + \beta_{j+1}, \quad i + j + 1 \leqslant n.$$

§4 正规阵与可正规化阵的特征值

4.1 正规阵与可正规化阵

Hermite 阵,斜 Hermite 阵(即 $A^H = -A$ 者)和酉阵，都是正规阵. 所以正规阵是一类重要的矩阵.

定义 4.1. 设 $A \in \mathbf{C}^{n\times n}$. 如果存在非奇异阵 R, 使得 $R^{-1}AR$ 是正规阵,则称 A 为可正规化阵.

容易看出，A是可正规化阵的必要与充分条件是存在非奇异阵 Q, 使得 $Q^{-1}AQ$ 为对角阵. 因此，可正规化阵就是可对角化阵.

所以,可正规化阵也可以如下定义. 设 $A \in \mathbf{C}^{n\times n}$. 如果 $\mathbf{C}^n$ 存在由 A的特征向量组成的基底,则称 A 为可正规化阵.

正规阵的特征值有许多重要的特点. 下述定理，说明了它的特点之一.

定理 4.1. 设 $A \in \mathbf{C}^{n\times n}$是正规阵，$\lambda(A) = \{\lambda_i\}$. 任取 $x \in \mathbf{C}^n$, $\|x\|_2 = 1$. 则

$$\min_{1\leqslant i\leqslant n} |\lambda_i - x^H A x| \leqslant [x^H A^H A x - |x^H A x|^2]^{\frac{1}{2}}$$

$$= \|Ax - x^H A x x\|_2. \tag{4.1}$$

证明:

考察 $\|(A - x^H A x I)x\|_2$.

由 A 是正规阵可知，存在酉阵 U，使得 $A = U\Lambda U^H$，其中 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 于是，一方面有

$$\begin{aligned}\|(A - \boldsymbol{x}^H A\boldsymbol{x}I)\boldsymbol{x}\|_2 &= \|U(\Lambda - \boldsymbol{x}^H A\boldsymbol{x}I)U^H\boldsymbol{x}\|_2 \\ &\geqslant \min_{1\leqslant i\leqslant n} |\lambda_i - \boldsymbol{x}^H A\boldsymbol{x}|;\end{aligned}$$

另一方面，有

$$\|(A - \boldsymbol{x}^H A\boldsymbol{x}I)\boldsymbol{x}\|_2 = (\boldsymbol{x}^H A^H A\boldsymbol{x} - |\boldsymbol{x}^H A\boldsymbol{x}|^2)^{\frac{1}{2}}.$$

所以(4.1)式成立. □

4.2 Hoffman-Wielandt 定理

定理 4.2 (Hoffman-Wielandt 定理)[104]. 设 A, $B \in \mathbf{C}^{n\times n}$ 皆为正规阵，则

$$e(A, B) \leqslant \|B - A\|_F. \tag{4.2}$$

其中 $e(A, B)$ 如(1.13)所定义.

为了证明定理 4.2，首先引证关于双随机阵的 Birkhoff 定理.

定义 4.2. 设 $S = (\sigma_{ij}) \in \mathbf{C}^{n\times n}$. 如果

$$\sum_{j=1}^{n} \sigma_{ij} = \sum_{i=1}^{n} \sigma_{ij} = 1, \ \sigma_{ij} \geqslant 0, 1 \leqslant i, \ j \leqslant n,$$

则称 S 为双随机阵.

排列方阵，即每行、每列有且仅有 1 个元素为 1，其余元素为零的方阵，显然是一类特殊的双随机阵；并且共有 $n!$ 个不同的 n 阶排列方阵.

定义 4.3 设 $T = (\tau_{ij}) \in \mathbf{C}^{n\times n}$. 如果 $\pi(1), \cdots, \pi(n)$ 是 $1, \cdots, n$ 的任一排列，则称集合

$$\tau_{1\pi(1)}, \cdots, \tau_{n\pi(n)}$$

是 T 的一个正则组.

定理 4.3 (Frobenius-König 定理). 设 $T = (\tau_{ij}) \in \mathbf{C}^{n\times n}$. 如果 T 的每个正则组都含有零元素，则 T 必有一子矩阵 $T_0 \in \mathbf{C}^{p\times q}$ 为零矩阵，并且 $p + q = n + 1$.

证明:

应用数学归纳法.

当 $n=1$ 时,结论显然成立. 假定结论对所有阶数小于 n 的方阵均成立,现考虑 $T\in\mathbf{C}^{n\times n}$.

如果 $T=0$, 则定理结论自然成立. 因此假定 $T\neq 0$, 并且因为行列置换并不影响结论,所以无妨假定 $\tau_{nn}\neq 0$.

令 T_1 表示 T 中左上角 $n-1$ 阶子矩阵. 由定理假设知, T_1 的每个正则组必含有零元素(否则, 添上 τ_{nn} 后, T 就有不含零元素的正则组,这与定理假设矛盾). 所以, 由归纳法假设, T_1 必有一子矩阵 $T_{10}\in\mathbf{C}^{p_1\times q_1}$ 为零矩阵, 并且 $p_1+q_1=n$. 无妨假设 T_{10} 位于 T_1 的左上角,这时

$$T=\left(\begin{array}{c:c} T_{10} & T_2 \\ \hdashline T_3 & \end{array}\right)\begin{array}{l}\}\,p_1\\ \}\,q_1.\end{array}\qquad \text{(列宽: } \underbrace{\quad}_{q_1}\ \underbrace{\quad}_{p_1}\text{)}$$

在 T 中, T_2 与 T_3 皆为方阵. 易知,在 T_2 与 T_3 之中, 必至少有一个满足定理的条件(否则,可以分别从 T_2 与 T_3 中取出不含零元素的正则组, 并在一起,便成了 T 的一个不含零元素的正则组, 这与定理假设矛盾). 假定 T_2 的每个正则组都含有零元素. 据归纳法假设, T_2 有一子矩阵 $T_{20}\in\mathbf{C}^{p_2\times q_2}$ 是零矩阵,并且 $p_2+q_2=p_1+1$. 无妨假定 T 具有下述形式 (因为进行行列置换并不影响定理的结论):

$$T=\left(\begin{array}{c:cc} T_{10} & \underbrace{T_{20}}_{q_2}\}\,p_2 & \\ \hdashline & & \\ \hdashline \underbrace{\qquad}_{q_1} & & \end{array}\right)\}\,p_1.$$

这时, T 的左上角是一个 p_2 行、(q_1+q_2) 列的零矩阵, 而

$$p_2+(q_1+q_2)=(p_2+q_2)+q_1=p_1+1+q_1=n+1.$$

因此定理得证. □

由定理 4.3 可得下述推论.

推论 4.1. 如果 T 为非负矩阵, 并且它的每行元素之和与每

列元素之和都等于同一个正数 τ，则 $\boldsymbol{T}$ 必含有由正元素组成的正则组.

证明:

用反证法.

如果 T 的每个正则组都含有零元素，则由定理 4.3，存在 T 的一个零子矩阵 $T_0\in \mathbb{C}^{p\times q}, p+q=n+1$. 于是在 T 中，T_0 所在的 p 行元素之和为 $p\tau$，T_0 所在的 q 列元素之和为 $q\tau$，而 $p\tau+q\tau=(n+1)\tau > T$ 的全部元素之和 $n\tau$，显然矛盾. □

利用推论 4.1 可证

定理 4.4 (Birkhoff 定理). 所有 n 阶双随机阵的集合，是所有 n 阶排列方阵的凸包. 即任一 n 阶双随机阵 S，必可表示成 n 阶排列方阵 $P_i(i=1,\cdots,n!)$ 的凸组合:

$$S=\sum_{i=1}^{n!}\sigma_i P_i,\quad \sum_{i=1}^{n!}\sigma_i=1,\ \sigma_i\geqslant 0,\ i=1,\cdots,n!. \tag{4.3}$$

证明:

首先指出，(4.3)所示的非负阵 S 是双随机阵. 这是因为，$\sigma_i P_i$ 的每行元素之和及每列元素之和都是 σ_i，所以 S 的每行元素之和及每列元素之和都是 $\sum_{i=1}^{n!}\sigma_i=1$.

现在证明任一 n 阶双随机阵必可表示成(4.3)的形式. 根据推论 4.1，S 必含有由正元素组成的正则组 $\sigma_{1i_1},\cdots,\sigma_{ni_n}$. 令 $\sigma_1=\min\limits_{1\leqslant k\leqslant n}\sigma_{ki_k}>0$，并且令 P_1 表示在 $(1,i_1),\cdots,(n,i_n)$ 位置为 1 的排列阵，然后考虑矩阵 $S_1=S-\sigma_1 P_1$.

显然 S_1 有三个特点: ①. 它是非负阵，②. 它的每行元素之和及每列元素之和都等于 $1-\sigma_1\geqslant 0$, ③. S_1 比 S 至少多一个零元素.

如果 $1-\sigma_1=0$，则 $S_1=0$，即 $S=\sigma_1 P_1$，定理得证. 如果 $1-\sigma_1>0$，则据推论 4.1，S_1 必含有由正元素组成的正则组. 重复前述讨论，得到矩阵 $S_2=S-\sigma_1P_1-\sigma_2P_2$. S_2 有三个特

点：①.它是非负阵，②.它的每行元素之和及每列元素之和都等于 $1-\sigma_1-\sigma_2\geqslant 0$，③.$S_2$ 比 S_1 至少多一个零元素，因而比 S 至少多两个零元素.

如此重复进行，每一步都得到一个新的非负阵，这个新的非负阵的每行元素之和及每列元素之和都等于同一个非负数，而且它比前一个非负阵至少多出一个零元素.

到了第 n^2-n-1 步，这时 S_{n^2-n-1} 至少含有 n^2-n-1 个零元素，即至多含有 $n+1$ 个非零元素，而且它的每行元素之和与每列元素之和应该相等. 由此可以断言：$S_{n^2-n-1}=0$ 或者 $S_{n^2-n-1}=\sigma_{n^2-n}P_{n^2-n}$. 因为，容易看出，如果 S_{n^2-n-1} 的某一行或某一列都是零元素，则应有 $S_{n^2-n-1}=0$. 如果 S_{n^2-n-1} 的每一行、每一列都有一个非零元素，同时它的每行元素之和及每列元素之和相等，则它只可能有 n 个非零元素(即这时不存在第 $n+1$ 个非零元素)，并且有 $S_{n^2-n-1}=\sigma_{n^2-n}P_{n^2-n}$.

所以最终得到 $S_m=S-\sigma_1P_1-\cdots-\sigma_mP_m=0$，其中 $m\leqslant n^2-n$，即

$$S=\sum_{i=1}^{m}\sigma_iP_i,\quad m\leqslant n^2-n\leqslant n!,$$

并且 $\{\sigma_i\}$ 满足 $\sum_{i=1}^{m}\sigma_i=1$，$\sigma_i>0$. □

Birkhoff 定理有许多应用. 下面就用它来证明 Hoffman-Wielandt 定理.

定理 4.2 的证明:

首先对正规阵 A 和 B 进行酉-对角分解:

$$A=U\Lambda U^H,\quad B=V\Omega V^H,\tag{4.4}$$

其中 U 与 V 为酉阵，$\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，$\Omega=\mathrm{diag}(\mu_1,\cdots,\mu_n)$. 利用 (4.4)，可得

$$\|B-A\|_F^2=\mathrm{tr}[(V\Omega V^H-U\Lambda U^H)(V\Omega V^H-U\Lambda U^H)^H]$$

$$=\sum_{i=1}^{n}|\lambda_i|^2+\sum_{i=1}^{n}|\mu_i|^2$$

$$-\operatorname{tr}(\Lambda U^H V \bar{\Omega} V^H U + \bar{\Lambda} U^H V \Omega V^H U)$$

$$= \sum_{i=1}^{n} |\lambda_i|^2 + \sum_{i=1}^{n} |\mu_i|^2 - g(W), \tag{4.5}$$

其中

$$W = U^H V = (\omega_{ij}) \tag{4.6}$$

为酉阵，

$$\left.\begin{aligned} g(W) &= \sum_{i,j=1}^{n} \theta_{ij} |\omega_{ij}|^2, \\ \theta_{ij} &= \lambda_i \bar{\mu}_j + \bar{\lambda}_i \mu_j, \ 1 \leqslant i, j \leqslant n. \end{aligned}\right\} \tag{4.7}$$

令

$$|\omega_{ij}|^2 = \sigma_{ij}, \ 1 \leqslant i, j \leqslant n, \tag{4.8}$$

则 $S = (\sigma_{ij})$ 显然是双随机阵，并且(4.7)可写成

$$g(W) \equiv f(S) = \sum_{i,j=1}^{n} \theta_{ij} \sigma_{ij} \tag{4.9}$$

用 $\mathscr{S}_n$ 表示 n 阶双随机阵的全体，用 $\mathscr{U}_n$ 表示 n 阶酉阵的全体，于是由(4.6)—(4.9)可知

$$\max_{W \in \mathscr{U}_n} g(W) \leqslant \max_{S \in \mathscr{S}_n} f(S). \tag{4.10}$$

注意到 $f(S)$ 是 $\mathscr{S}_n$ 上的线性函数，并且根据 Birkhoff 定理，S 可以表示成 n 阶排列方阵 P_k 的凸组合，即如 (4.3) 式所示。代入 $f(S)$，便得到

$$f(S) = f\left(\sum_{k=1}^{n!} \sigma_k P_k\right) = \sum_{k=1}^{n!} \sigma_k f(P_k),$$

其中 $\sigma_k \geqslant 0, k = 1, \cdots, n!, \sum_{k=1}^{n} \sigma_k = 1$. 设排列方阵

$$P = (\delta_{i\pi(i)}), \quad \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

使得

$$\max_{1 \leqslant k \leqslant n!} f(P_k) = f(P),$$

则有

$$f(S) \leqslant \sum_{k=1}^{n_1} \sigma_k f(P) = f(P). \tag{4.11}$$

代入(4.10)和(4.9)，得到

$$g(W) \leqslant \sum_{i=1}^{n} \theta_{i\pi(i)},$$

其中 $\pi(1),\cdots,\pi(n)$ 是 $1,\cdots,n$ 的某一个排列.

另一方面，取 $W = P$，有

$$g(P) = \sum_{i=1}^{n} \theta_{i\pi(i)};$$

所以

$$\max_{W\in\mathscr{U}_n} g(W) = \sum_{i=1}^{n} \theta_{i\pi(i)} = \sum_{i=1}^{n} (\lambda_i \bar{\mu}_{\pi(i)} + \bar{\lambda}_i \mu_{\pi(i)}).$$

代入(4.5)，便得出

$$\|B - A\|_F^2 \geqslant \sum_{i=1}^{n} |\lambda_i - \mu_{\pi(i)}|^2,$$

由此立即得到不等式 (4.2). □

注 4.1 同理可证，存在 $1,\cdots,n$ 的一个排列 $\rho(1),\cdots,\rho(n)$，使得

$$\|B - A\|_F \leqslant \sqrt{\sum_{i=1}^{n} |\lambda_i - \mu_{\rho(i)}|^2}.$$

注 4.2 如果 A 与 B 不同时为正规阵，则定理 4.2 的结论不一定成立. 例如

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}.$$

A 正规，但 B 不正规. 有 $\lambda(A) = \{0,4\}$，$\lambda(B) = \{0,0\}$，以及

$$e^2(A,B) = 16 > 12 = \|B - A\|_F^2.$$

注 4.3 当 B 为正规阵而 A 为 Hermite 阵时，可证(见本节习

题 1): 如果将 $\lambda(A)=\{\alpha_k\}$ 与 $\lambda(B)=\{\beta_k+i\gamma_k\}$ 分别排序为

$$\alpha_1\geqslant\cdots\geqslant\alpha_n,\ \beta_1\geqslant\cdots\geqslant\beta_n,$$

则

$$e(A,B)=\sqrt{\sum_{k=1}^{n}|(\beta_k+i\gamma_k)-\alpha_k|^2}\leqslant\|B-A\|_F. \tag{4.12}$$

注 4.4 Birkhoff 定理的几何意义是，所有 n 阶双随机阵组成的集合 $\mathscr{S}_n$ 是以 n 阶排列方阵为顶点的凸多面体. 因为在一个凸多面体上定义的线性函数必在该凸多面体的顶点上达到极大值和极小值，所以(4.11)也可直接由此得证.

注 4.5 当 $A, B\in\mathbb{C}^{n\times n}$ 都是正规阵时，是否有类似于定理 3.6 的结论，即是否有

$$v(A,B)\leqslant\|B-A\|_2,$$

这是一个尚未解决的问题(请参考[58]).

4.3 Bauer-Fike 定理

定理 4.5 (Bauer-Fike 定理)[48]. 设 $A, B\in\mathbb{C}^{n\times n}$，其中 A 为可正规化阵: $A=Q^{-1}\Lambda Q$，Q 为非奇异阵，$\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$. 则

$$s_A(B)\leqslant\|Q\|_2\|Q^{-1}\|_2\|B-A\|_2. \tag{4.13}$$

证明:

任取 $\mu\in\lambda(B)$. 显然只需考虑 $\mu\notin\lambda(A)$ 的情形，即 $\mu I-\Lambda$ 非奇异.

设 $\boldsymbol{x}$ 是 B 属于 μ 的特征向量. 于是有

$$(B-A)\boldsymbol{x}=(\mu I-A)\boldsymbol{x}=Q^{-1}(\mu I-\Lambda)Q\boldsymbol{x},$$

即

$$\boldsymbol{x}=Q^{-1}(\mu I-\Lambda)^{-1}Q(B-A)\boldsymbol{x}.$$

因而对于任一相容的范数 $\|\ \ \|$，有

$$\begin{aligned}1&\leqslant\|Q^{-1}(\mu I-\Lambda)^{-1}Q(B-A)\|\\&\leqslant\|Q\|\|Q^{-1}\|\|(\mu I-\Lambda)^{-1}\|\|B-A\|,\end{aligned}$$

即

$$\|(\mu I - A)^{-1}\|^{-1} \leqslant \|Q\|\|Q^{-1}\|\|B - A\|. \tag{4.14}$$

在(4.14)中取谱范数 $\|\ \|_2$，便得出 (4.13)．□

推论 4.2． 设 $A, B \in \mathbf{C}^{n\times n}$，其中 A 为正规阵．则

$$S_A(B) \leqslant \|B - A\|_2. \tag{4.15}$$

注 4.6． (4.13)右端 $\|B - A\|_2$ 前的系数

$$\kappa(Q) = \|Q\|_2\|Q^{-1}\|_2, \tag{4.16}$$

叫做 A 的谱条件数．对于可正规化阵 A，令

$$\mathscr{Q}_A = \{Q \in \mathbf{C}_n^{n\times n}: Q^{-1}AQ = \Lambda\},$$

其中 Λ 为对角阵．有时也把

$$\nu(A) = \inf_{Q\in\mathscr{Q}_A} \|Q\|_2\|Q^{-1}\|_2 \tag{4.17}$$

叫做 A 的谱条件数．显然 $\kappa(Q) \geqslant 1$，$\nu(A) \geqslant 1$．而 (4.15) 表明，正规阵的谱条件数为 1．

4.4 Hermite 阵的任意扰动

定理 4.6 (Kahan, [115])．设 $A \in \mathbf{C}^{n\times n}$ 为 Hermite 阵，其特征值为 $\alpha_1 \geqslant \cdots \geqslant \alpha_n$．又设 $B = A + E$，$\lambda(B) = \{\beta_k + i\gamma_k\}$，$\beta_1 \geqslant \cdots \geqslant \beta_n$．令

$$E_x = \frac{E + E^H}{2}, \quad E_y = \frac{E - E^H}{2i} \tag{4.18}$$

和

$$\mathscr{D}_k = \{\beta + i\gamma \in \mathbf{C}: |\beta + i\gamma - \alpha_k| \leqslant \|E\|_2,\ |\gamma| \leqslant \|E_y\|_2\}, \tag{4.19}$$

则

$$\lambda(B) \subset \bigcup_{k=1}^{n} \mathscr{D}_k. \tag{4.20}$$

证明:

由定理 4.5 知，对于任一 $\beta + i\gamma \in \lambda(B)$，必存在某一 α_k．使得

$$|\beta + i\gamma - \alpha_k| \leqslant \|E\|_2.$$

以下证明: $|\gamma| \leqslant \|E_y\|_2$.

设 $\boldsymbol{x}$ 是 B 属于 $\beta + i\gamma$ 的单位特征向量,即

$$B\boldsymbol{x} = (\beta + i\gamma)\boldsymbol{x}, \quad \|\boldsymbol{x}\|_2 = 1.$$

则

$$\boldsymbol{x}^H B \boldsymbol{x} = \beta + i\gamma, \quad \boldsymbol{x}^H B^H \boldsymbol{x} = \beta - i\gamma,$$

因而

$$\boldsymbol{x}^H E_y \boldsymbol{x} = \frac{\boldsymbol{x}^H (B - B^H) \boldsymbol{x}}{2i} = \gamma,$$

由此得出 $|\gamma| \leqslant \|E_y\|_2$. □

定理 4.6 的结论,如图 3-2 所示.

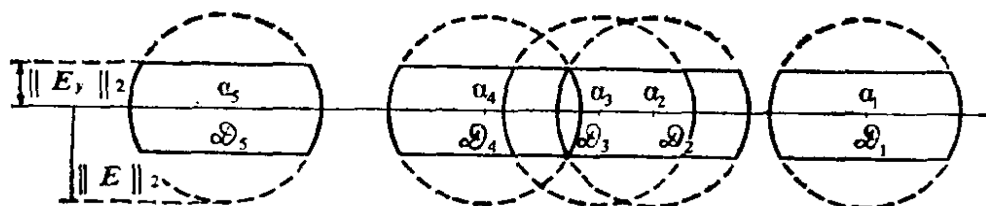


图 3-2

图 3-2 表示: $n = 5$, $\beta_1 + i\gamma_1 \in \mathscr{D}_1$, $\beta_5 + i\gamma_5 \in \mathscr{D}_5$, 其它三个点 $\beta_k + i\gamma_k \in \mathscr{D}_2 \cup \mathscr{D}_3 \cup \mathscr{D}_4$, $k = 2,3,4$.

定理 4.7 (Kahan,[115]). 在定理 4.6 的假设条件下,有

$$\sqrt{\sum_{i=1}^{n} \gamma_k^2} \leqslant \|E_y\|_F \tag{4.21}$$

和

$$\sqrt{\sum_{k=1}^{n} (\beta_k - \alpha_k)^2} \leqslant \|E_x\|_F + \sqrt{\|E_y\|_F^2 - \sum_{k=1}^{n} \gamma_k^2}. \tag{4.22}$$

证明:

利用 Schur 分解,无妨假定

$$B \equiv A + E = \Omega + i\Gamma + R, \tag{4.23}$$

其中 $\Omega = \mathrm{diag}(\beta_1, \cdots, \beta_n)$, $\Gamma = \mathrm{diag}(\gamma_1, \cdots, \gamma_n)$, R 为严格上三角阵. 于是

$$A + E_x = \Omega + \frac{R + R^H}{2},\ E_y = \Gamma + \frac{R - R^H}{2i}. \tag{4.24}$$

从而得到

$$\|E_y\|_F^2 = \operatorname{tr}\left(\Gamma + \frac{R - R^H}{2i}\right)\left(\Gamma + \frac{R - R^H}{2i}\right)$$

$$= \|\Gamma\|_F^2 + \frac{1}{2}\|R\|_F^2 \geqslant \|\Gamma\|_F^2 = \sum_{k=1}^{n} \gamma_k^2. \tag{4.25}$$

另一方面,对于 A 与 Ω 应用 Hoffman-Wielandt 定理,并利用(4.24)和(4.25),可得

$$\sqrt{\sum_{k=1}^{n}(\beta_k - \alpha_k)^2} \leqslant \|\Omega - A\|_F = \left\|E_x - \frac{R + R^H}{2}\right\|_F$$

$$\leqslant \|E_x\|_F + \left\|\frac{R + R^H}{2}\right\|_F = \|E_x\|_F + \frac{1}{\sqrt{2}}\|R\|_F$$

$$= \|E_x\|_F + \sqrt{\|E_y\|_F^2 - \|\Gamma\|_F^2} = \|E_x\|_F + \sqrt{\|E_y\|_F^2 - \sum_{k=1}^{n}\gamma_k^2}.$$

□

注意到,当 $B = A + E$ 是正规阵时,(4.23)中的 $R = 0$. 所以,由(4.21)、(4.22)和 (4.25) 立即导出

推论 4.3. 在定理 4.6 的假设条件下,如果再假设 $B = A+E$ 是正规阵,则有

$$\sqrt{\sum_{k=1}^{n} \gamma_k^2} \leqslant \|E_y\|_F, \sqrt{\sum_{k=1}^{n}(\beta_k - \alpha_k)^2} \leqslant \|E_x\|_F. \tag{4.26}$$

值得指出的是,如果利用 Hoffman-Wielandt 定理到 A 与 $B = A + E$, 在推论4.3的条件下,只能得出

$$\sqrt{\sum_{k=1}^{n} |(\beta_k + i\gamma_k) - \alpha_k|^2} \leqslant \|E\|_F. \tag{4.27}$$

由

$$|(\beta_k + i\gamma_k) - \alpha_k|^2 = (\beta_k - \alpha_k)^2 + \gamma_k^2$$

和

$$\|E\|_F^2 = \|E_x\|_F^2 + \|E_y\|_F^2$$

可知，从(4.26)能够导出(4.27). 所以，推论 4.3 的结论，要比利用 Hoffman-Wielandt 定理得到的结论更强.

在定理 4.6 的条件下，有下述不等式:

$$\begin{aligned}
&\sum_{k=1}^{n}(\beta_k-\alpha_k)^2+2\sum_{k=1}^{n}\gamma_k^2 \\
&\leqslant\left(\|E_x\|_F+\sqrt{\|E_y\|_F^2-\sum_{k=1}^{n}\gamma_k^2}\right)^2+2\sum_{k=1}^{n}\gamma_k^2 \\
&=2\|E_x\|_F^2+2\|E_y\|_F^2 \\
&\qquad-\left(\|E_x\|_F-\sqrt{\|E_y\|_F^2-\sum_{k=1}^{n}\gamma_k^2}\right)^2 \\
&\leqslant 2(\|E_x\|_F^2+\|E_y\|_F^2)=2\|E\|_F^2,
\end{aligned}\tag{4.28}$$

因此可得

推论 4.4. 在定理 4.6 的假设条件下，有

$$\begin{aligned}
&\sqrt{\sum_{k=1}^{n}|(\beta_k+i\gamma_k)-\alpha_k|^2} \\
&\leqslant\sqrt{\sum_{k=1}^{n}(\beta_k-\alpha_k)^2+2\sum_{k=1}^{n}\gamma_k^2}\leqslant\sqrt{2}\|E\|_F.
\end{aligned}\tag{4.29}$$

注 4.7. 现举一例，说明估计式(4.21)与(4.22)，以及(4.28)与(4.29).

$$A=\begin{pmatrix}0&1\\1&0\end{pmatrix},\ E=\begin{pmatrix}i\gamma_1&0\\-1&i\gamma_2\end{pmatrix},\quad B\equiv A+E=\begin{pmatrix}i\gamma_1&1\\0&i\gamma_2\end{pmatrix}.$$

有

$$\alpha_1=1,\ \alpha_2=-1,\ \beta_1=\beta_2=0;\ \|E_x\|_F^2=\frac{1}{2},$$

$$\|E_y\|_F^2=\frac{1}{2}+\gamma_1^2+\gamma_2^2,\ \|E\|_F^2=1+\gamma_1^2+\gamma_2^2.$$

于是得到

$$\sqrt{\sum_{k=1}^{2}\gamma_k^2} < \|E_y\|_F = \sqrt{\frac{1}{2}+\sum_{k=1}^{2}\gamma_k^2},$$

$$\sqrt{2} = \sqrt{\sum_{k=1}^{2}(\beta_k-\alpha_k)^2} = \|E_x\|_F + \sqrt{\|E_y\|_F^2 - \sum_{k=1}^{2}\gamma_k^2}$$

和

$$\sqrt{2+\sum_{k=1}^{2}\gamma_k^2} = \sqrt{\sum_{k=1}^{2}|(\beta_k+i\gamma_k)-\alpha_k|^2}$$

$$< \sqrt{\sum_{k=1}^{2}(\beta_k-\alpha_h)^2 + 2\sum_{k=1}^{2}\gamma_k^2} = \sqrt{2}\|E\|_F$$

$$= \sqrt{2\left(1+\sum_{k=1}^{2}\gamma_k^2\right)}.$$

这些结果表明,上例使(4.22)和(4.28)达到了等式.

对于 Hermite 阵 A, 如果知道 $B=A+E$ 相似于一个 Hermite 阵,则可应用下述定理的估计.

定理4.8 (Kahan,[115]). 设 $A\in \mathbf{C}^{n\times n}$ 为 Hermite 阵,其特征值为 $\alpha_1\geqslant\cdots\geqslant\alpha_n$. 又设对于 $B=A+E$, 存在非奇异阵 Q, 使得 $B_Q=Q(A+E)Q^{-1}$ 是一 Hermite 阵. 记 B 的特征值(即 B_Q 的特征值)为 $\beta_1\geqslant\cdots\geqslant\beta_n$. 则有

$$|\beta_k-\alpha_k| \leqslant \|Q\|_2\|Q^{-1}\|_2\|E\|_2, k=1,\cdots,n. \qquad (4.30)$$

证明:

首先将 Q 进行极分解: $Q=UH$,其中 U 为酉阵, $H=(Q^HQ)^{\frac{1}{2}}>0$. 显然有 $\|H\|_2=\|Q\|_2$ 和 $\|H^{-1}\|_2=\|Q^{-1}\|_2$. 记

$$\hat{B}=U^HB_QU=H(A+E)H^{-1},$$

则 $\hat{B}$ 仍为 Hermite 阵, 并且 $\lambda(\hat{B})=\lambda(B_Q)=\lambda(B)$. 据定理 3.6,

$$|\beta_k-\alpha_k| \leqslant \|\hat{B}-A\|_2, k=1,\cdots,n.$$

所以下面只需证明

$$\|\hat{B}-A\|_2 \leqslant \|Q\|_2\|Q^{-1}\|_2\|E\|_2. \qquad (4.31)$$

设 $\boldsymbol{x}$ 是 $\hat{B}-A$ 属于绝对值最大的特征值的单位特征向量,

即

$$(\hat{B}-A)\boldsymbol{x}=\pm\|\hat{B}-A\|_2\boldsymbol{x}, \tag{4.32}$$

并且 $\|\boldsymbol{x}\|_2=1$. 于是有

$$\begin{aligned}
\|Q\|_2\|E\|_2 &= \|H\|_2\|Q^{-1}B_QQ-A\|_2 \\
&= \|H\|_2\|Q^{-1}U\hat{B}U^HQ-A\|_2 \\
&= \|H\|_2\|H^{-1}\hat{B}H-A\|_2 \\
&= \|H\|_2\|H^{-1}(\hat{B}H-HA)\|_2 \\
&\geqslant \|\hat{B}H-HA\|_2 \\
&\geqslant |\boldsymbol{x}^H(\hat{B}H-HA)\boldsymbol{x}| \\
&= |\boldsymbol{x}^H(\hat{B}H-H\hat{B})\boldsymbol{x}+\boldsymbol{x}^HH(\hat{B}-A)\boldsymbol{x}| \\
&= |\boldsymbol{x}^H(\hat{B}H-H\hat{B})\boldsymbol{x}\pm\|\hat{B}-A\|_2\boldsymbol{x}^HH\boldsymbol{x}| \qquad (\text{据}(4.32)) \\
&\geqslant \|\hat{B}-A\|_2\boldsymbol{x}^HH\boldsymbol{x} \\
&\geqslant \|\hat{B}-A\|_2/\|H^{-1}\|_2 \\
&= \frac{\|\hat{B}-A\|_2}{\|Q^{-1}\|_2}.
\end{aligned}$$

由此立即得到(4.31). □

习题

1. 设 $B\in\mathbf{C}^{n\times n}$ 为正规阵, $A\in\mathbf{C}^{n\times n}$ 为 Hermite 阵, $\lambda(A)=\{\alpha_k\}$ 与 $\lambda(B)=\{\beta_k+i\gamma_k\}$ 分别排序为 $\alpha_1\geqslant\cdots\geqslant\alpha_n$ 和 $\beta_1\geqslant\cdots\geqslant\beta_n$. 则

$$e(A,B)\triangleq\sqrt{\sum_{k=1}^{n}|(\beta_k+i\gamma_k)-\alpha_k|^2}\leqslant\|B-A\|_F.$$

2. 设 $A\in\mathbf{C}^{n\times n}$ 为正规阵, $\lambda(A)=\{\lambda_i\}$, $D=\mathrm{diag}(\delta_1,\cdots,\delta_n)$, $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T\in\mathbf{C}^n$ 满足 $A\boldsymbol{x}=D\boldsymbol{x}\neq0$. 又设

$$\delta_1,\cdots,\delta_n\in\mathscr{D}_r(z_0)=\{z\in\mathbf{C}:|z-z_0|\leqslant r\}.$$

试证: 必有 $\lambda_{i_0}\in\lambda(A)$, 使得 $\lambda_{i_0}\in\mathscr{D}_r(z_0)$.

3. 设 $A\in\mathbf{C}^{n\times n}$ 为正规阵, $\lambda(A)=\{\alpha_i\}$ 满足 $|\alpha_1|\geqslant\cdots\geqslant|\alpha_n|$. 又设 Q 为酉阵, $\lambda(AQ)=\{\lambda_i\}$. 试证:

$$|\alpha_1| \geqslant |\lambda_i| \geqslant |\alpha_n|, \quad i = 1, \cdots, n.$$

并由此证明：若 $U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}$ 为酉阵，$\mu \in \mathbb{C}$ 满足

$$\det \begin{pmatrix} U_{11} - \mu I & U_{12} \\ U_{21} & U_{22} \end{pmatrix} = 0,$$

则 $|\mu| \geqslant 1$.

4. 在定理 4.8 的条件下，试证

$$\sqrt{\sum_{k=1}^{n} (\beta_k - \alpha_k)^2} \leqslant \|Q\|_2 \|Q^{-1}\|_2 \|E\|_F.$$

§5 一般方阵的特征值

5.1 推广的 Bauer-Fike 定理

定理 5.1 (Kahan, Parlett, 蒋, [116]). 设 $A = Q^{-1} J Q \in \mathbb{C}^{n \times n}$, J 是 A 的 Jordan 标准形. 又设 $B = A + E$. 则对于任一 $\mu \in \lambda(B)$, 必有 $\lambda \in \lambda(A)$, 使得

$$\frac{|\lambda - \mu|^m}{(1 + |\lambda - \mu|)^{m-1}} \leqslant \|QEQ^{-1}\|_2, \tag{5.1}$$

其中 m 是 J 中属于 λ 的最大 Jordan 块的阶数.

证明:

若 $\mu \in \lambda(A)$, 则(5.1)显然成立. 现设 $\mu \notin \lambda(A)$, 这时 $A + E - \mu I$ 奇异, 但 $J - \mu I$ 非奇异. 将 A 的 Jordan 分解代入, 知

$$\begin{aligned} & Q^{-1}[(J - \mu I) + QEQ^{-1}]Q \\ & = Q^{-1}(J - \mu I)[I + (J - \mu I)^{-1}QEQ^{-1}]Q \end{aligned}$$

为奇异阵, 因而 $I + (J - \mu I)^{-1}QEQ^{-1}$ 必为奇异阵. 于是对于 $\mathbb{C}^{n \times n}$ 上任一种相容的范数 $\| \quad \|$, 有

$$\|(J - \mu I)^{-1}QEQ^{-1}\| \geqslant 1,$$

从而

$$\|(J - \mu I)^{-1}\|^{-1} \leqslant \|QEQ^{-1}\|. \tag{5.2}$$

以下利用谱范数 $\| \quad \|_2$ 估计 $\|(J - \mu I)^{-1}\|_2^{-1}$ (即 $J - \mu I$ 的

最小奇异值)的下界.

考虑 $J-\mu I$ 的任一个 k 阶 Jordan 块

$$J_1=\begin{pmatrix}\delta & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & \delta\end{pmatrix},$$

其中 $\delta=\lambda-\mu$, $\lambda\in\lambda(A)$, $\lambda\neq\mu$. 令

$$T_1=J_1J_1^H=\begin{pmatrix}1+|\delta|^2 & \bar{\delta} & & \\ \delta & 1+|\delta|^2 & \ddots & \\ & \ddots & \ddots & \ddots \\ & & \ddots & 1+|\delta|^2 & \bar{\delta}\\ & & & \delta & |\delta|^2\end{pmatrix}.$$

T_1 显然是正定阵. 设 T_1 的特征值为 $\tau_1\geqslant\cdots\geqslant\tau_k(>0)$, 则据 Gerschgorin 定理, $\tau_i\leqslant 1+|\delta|^2+2|\delta|=(1+|\delta|)^2$, $i=1,\cdots,k$. 于是

$$\tau_k=\frac{\tau_1\cdots\tau_{k-1}\tau_k}{\tau_1\cdots\tau_{k-1}}=\frac{\det T_1}{\tau_1\cdots\tau_{k-1}}\geqslant\frac{|\delta|^{2k}}{(1+|\delta|)^{2(k-1)}}. \tag{5.3}$$

注意到函数 $\dfrac{\xi^k}{(1+\xi)^{k-1}}$ $(\xi>0)$ 随自然数 k 的增大而减小,所以必有一个形如 $\dfrac{|\lambda-\mu|^m}{1+|\lambda-\mu|^{m-1}}$ 的数作为 $\|(J-\mu I)^{-1}\|_2^{-1}$ 的下界,其中 λ 是 A 的某一个特征值,m 是 J 中属于 λ 的 Jordan 块的最大阶数. 代入(5.2)式左端,即得定理. □

推论 5.1. 设 $A\in\mathbf{C}^{n\times n}$ 是可正规化阵,即 $A=Q^{-1}\Lambda Q$, Λ 为对角阵. 又设 $B=A+E$. 则

$$s_A(B)\leqslant\|QEQ^{-1}\|_2. \tag{5.4}$$

因为 $\|QEQ^{-1}\|_2\leqslant\|Q\|_2\|Q^{-1}\|_2\|E\|_2$, 所以(5.4)比 Bauer-Fike 定理的结论(见(4.13))更强.

定理 5.1 已断言, 对于任一 $\mu\in\lambda(B)$, 必有 $\lambda\in\lambda(A)$, 使(5.1)式成立. 再注意到当 $\xi>0$ 时,函数 $\xi^m/(1+\xi)^{m-1}$ 随 ξ 减小而减小,而且也随 m 增大而减小. 因此,如果 A 的最大 Jordan 块

的阶数为 l，则有

$$\frac{[s_A(B)]^l}{[1+s_A(B)]^{l-1}} \leqslant \|QEQ^{-1}\|_2. \tag{5.5}$$

令

$$\eta = p_l(\xi) \equiv \frac{\xi^l}{(1+\xi)^{l-1}}, \xi \geqslant 0. \tag{5.6}$$

易知 $p_l(\xi)$ 在 $\xi \geqslant 0$ 上严格单调递增,因而对任一 $\eta \geqslant 0$，$p_l(\xi) = \eta$ 存在唯一的非负解，记为 $\xi = q_l(\eta)$. 即 $q_l(\eta)$ 是 $p_l(\xi)$ 在 $\xi \geqslant 0$ 上的反函数. 所以从(5.5)立即得出

$$s_A(B) \leqslant q_l(\|QEQ^{-1}\|_2). \tag{5.7}$$

再注意到 $q_l(\eta)$ 的下述性质：当 $\eta \leqslant \frac{1}{2^{l-1}}$ 时，$\xi \leqslant 1$，从而

$$\frac{q_l(\eta)}{\eta^{\frac{1}{l}}} = (1+\xi)^{1-\frac{1}{l}} \leqslant 2^{1-\frac{1}{l}},$$

即

$$q_l(\eta) \leqslant 2^{1-\frac{1}{l}} \eta^{\frac{1}{l}} \quad \left(\eta \leqslant \frac{1}{2^{l-1}}\right). \tag{5.8}$$

因此可得

推论 5.2. 在定理 5.1 的假设条件下,如果 $l(\geqslant 2)$ 是 A 的最大 Jordan 块的阶数,则当

$$\|QEQ^{-1}\|_2 \leqslant \frac{1}{2^{l-1}}$$

时,有

$$s_A(B) \leqslant 2^{1-\frac{1}{l}} \left\|QEQ^{-1}\right\|_2^{\frac{1}{l}} \tag{5.9}$$

$$\leqslant 2^{1-\frac{1}{l}} (\|Q\|_2 \|Q^{-1}\|_2)^{\frac{1}{l}} \|B-A\|_2^{\frac{1}{l}}. \tag{5.10}$$

5.2 Henrici 定理

Henrici [102] 利用矩阵的正规性偏离度,对任一矩阵特征值的扰动上界给出了估计.

定义 5.1. 设 $A \in \mathbf{C}^{n\times n}$，$\nu$ 是 $\mathbf{C}^{n\times n}$ 上的任一种范数。令

$\mathscr{M}_A = \{M : A$ 的 Schur 上三角形式中的严格上三角阵$\}$.
则

$$\Delta_\nu(A) = \inf_{M \in \mathscr{M}_A} \nu(M) \tag{5.11}$$

叫做 A 对于范数 ν 的正规性偏离度.

特别地,A 对于范数$\| \ \|_F$ 的正规性偏离度记作 $\Delta_F(A)$, A 对于范数$\| \ \|^2$的正规性偏离度记作 $\Delta_2(A)$.

推论 5.3. 设 $A \in \mathbf{C}^{n\times n}$, $\lambda(A) = \{\lambda_i\}$. 则

$$\Delta_F(A) = \sqrt{\|A\|_F^2 - \sum_{i=1}^n |\lambda_i|^2}. \tag{5.12}$$

证明:

设 T 是 A 的任一 Schur 上三角形式,即

$$U^H A U = T \equiv \Lambda + M,$$

其中 U 为酉阵, $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, M 为严格上三角阵. 根据 $\Delta_F(A)$ 的定义(见(5.11)),由

$$\|A\|_F^2 = \|T\|_F^2 = \|\Lambda\|_F^2 + \|M\|_F^2 = \sum_{i=1}^n |\lambda_i|^2 + \|M\|_F^2$$

立即得出 (5.12). □

由(5.12)可知,A 是正规阵的必要与充分条件是 $\Delta_F(A) = 0$. 再利用第二章定理 2.1,得到

推论 5.4. A 是正规阵的必要与充分条件是对于任一种矩阵范数 ν, 有 $\Delta_\nu(A) = 0$.

对于实变数 $\xi \geqslant 0$, 定义函数 f:

$$f(\xi) = \xi + \xi^2 + \cdots + \xi^n. \tag{5.13}$$

$f(\xi)$ 的反函数用 $g(\eta)$ 表示,即 $g(\eta)$是

$$g + g^2 + \cdots + g^n = \eta \ (\eta \geqslant 0) \tag{5.14}$$

的唯一非负解.

容易证明 f 与 g 有下列性质(本节习题 1)

(i) $f(\xi), g(\eta)$ 与 $\dfrac{f(\xi)}{\xi} = \dfrac{\eta}{g(\eta)}$ 皆为单调递增函数,

(ii) $\lim\limits_{\xi\to+0}\dfrac{f(\xi)}{\xi}=1,\quad \lim\limits_{\eta\to+0}\dfrac{g(\eta)}{\eta}=1$ (5.15)

(iii) $\dfrac{\eta}{n}\leqslant g(\eta)\leqslant\eta, 0\leqslant\eta\leqslant n,\ g(n)=1,$

$$\left(\frac{\eta}{n}\right)^{\frac{1}{n}}\leqslant g(\eta)\leqslant\eta^{\frac{1}{n}},\ \eta\geqslant n, \tag{5.16}$$

(iv) $\dfrac{\eta^{\frac{1}{n}}}{g(\eta)}$ 单调递降，并且

$$\lim_{\eta\to\infty}\frac{\eta^n}{g(\eta)}=1.$$

引理 5.1. 设 $R=D-M\in\mathbb{C}^{n\times n}$ 为非奇异上三角阵，其中 $D=\mathrm{diag}(\delta_1,\cdots,\delta_n)$，$M$ 为严格上三角阵. 如果

$$m\equiv\|M\|_2\neq 0,\quad \|R^{-1}\|_2^{-1}\leqslant\varepsilon. \tag{5.17}$$

则

$$\min_{1\leqslant i\leqslant n}|\delta_i|\leqslant\frac{m}{g\left(\dfrac{m}{\varepsilon}\right)}, \tag{5.18}$$

其中 g 是 $f(\xi)=\xi+\cdots+\xi^n(\xi\geqslant 0)$ 的反函数.

证明:

首先利用 $(D^{-1}M)^n=0$ 展开

$$\begin{aligned}R^{-1}&=(D-M)^{-1}=(I-D^{-1}M)^{-1}D^{-1}\\&=[I+D^{-1}M+\cdots+(D^{-1}M)^{n-1}]D^{-1}.\end{aligned} \tag{5.19}$$

令

$$\frac{1}{\min\limits_{1\leqslant i\leqslant n}|\delta_i|}=p,$$

则由(5.17)和(5.19)得到

$$\begin{aligned}\frac{1}{\varepsilon}&\leqslant\|(D-M)^{-1}\|_2\leqslant p+mp^2+\cdots+m^{n-1}p^n\\&=\frac{1}{m}f(mp),\end{aligned}$$

即 $\frac{m}{\varepsilon} \leqslant f(mp)$. 因为 $f(\xi)$ 单调递增,所以 $g\left(\frac{m}{\varepsilon}\right) \leqslant mp$,

即

$$\frac{1}{p} \leqslant \frac{m}{g\left(\frac{m}{\varepsilon}\right)},$$

这正是 (5.18)式. □

在叙述 Henrici 定理之前,还需说明"强范数"这个概念.

定义 5.2. 设 ν 与 ν' 是 $\mathbf{C}^{n\times n}$ 上的任意两种范数. 如果

$$\nu(A) \geqslant \nu'(A),\ \forall A \in \mathbf{C}^{n\times n},$$

则称 ν 是 ν' 的强范数.

定理 5.2 (Henrici 定理)[102]. 设 A, $B \in \mathbf{C}^{n\times n}$, A 为非正规阵,且 $B \neq A$, 如果 ν 是谱范数 $\|\ \ \|_2$ 的任一种强范数,则

$$s_A(B) \leqslant \frac{\eta}{g(\eta)} \nu(B-A),\quad \eta = \frac{\Delta_\nu(A)}{\nu(B-A)}. \tag{5.20}$$

证明:

据 Schur 定理,

$$A = UTU^H,\ T = \Lambda + M,$$

其中 U 为酉阵, $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, $\{\lambda_i\} = \lambda(A)$, M 为严格上三角阵. 令

$$U^H BU = B_1,\ U^H(B-A)U = E_1,$$

则有

$$B_1 = \Lambda + M + E_1. \tag{5.21}$$

任取 $\mu \in \lambda(B)$. 如果 $\mu \in \lambda(A)$, 则 (5.20) 显然成立. 以下考虑 $\mu \notin \lambda(A)$.

由(5.21)知,

$$E_1 = B_1 - \Lambda - M = (B_1 - \mu I) + (\mu I - \Lambda - M),$$

其中 $\mu \in \lambda(B_1)$, $\mu I - \Lambda - M$ 为非奇异阵. 设 $\boldsymbol{x}$ 是 B_1 属于 μ 的特征向量, 即 $B_1\boldsymbol{x} = \mu\boldsymbol{x}$, 则

$$E_1\boldsymbol{x} = (\mu I - \Lambda - M)\boldsymbol{x},\ \boldsymbol{x} = (\mu I - \Lambda - M)^{-1}E_1\boldsymbol{x}.$$

利用谱范数$\|\ \ \|_2$，得到

$$1 \leqslant \|(\mu I - A - M)^{-1}E_1\|_2 \leqslant \|(\mu I - A - M)^{-1}\|_2\|E_1\|_2,$$
$$\|(\mu I - A - M)^{-1}\|_2^{-1} \leqslant \|E_1\|_2 = \|B - A\|_2.$$

于是根据引理 5.1，有

$$\min_{1\leqslant i\leqslant n}|\lambda_i - \mu| \leqslant \frac{\|M\|_2}{g\left(\frac{\|M\|_2}{\|B-A\|_2}\right)}. \tag{5.22}$$

设 ν 是谱范数的强范数，则根据函数 g 的单调递增性和不等式(5.22)，得到

$$\left.\begin{aligned} s_A(B) &\leqslant \frac{\|M\|_2}{g\left(\frac{\|M\|_2}{\|B-A\|_2}\right)} \leqslant \frac{\eta_1}{g(\eta_1)}\nu(B-A) \\ \eta_1 &= \frac{\|M\|_2}{\nu(B-A)}. \end{aligned}\right\} \tag{5.23}$$

令 $\eta_2 = \frac{\nu(M)}{\nu(B-A)}$. 显然 $0 < \eta_1 < \eta_2$. 所以根据 $\frac{\eta}{g(\eta)}$ 的单调递增性及 g 的连续性，从(5.23)立即导出(5.20). □

注 5.1. 当 A 为正规阵时，$\nu(A) = 0$. 因此可以利用(5.15)，从(5.20)得到

$$s_A(B) \leqslant \nu(B-A),$$

其中 ν 是谱范数的任一种强范数. 所以，Bauer-Fike 定理的推论 4.2，也可以看作 Henrici 定理的一个特例.

注意到函数 $g(\eta)$ 的性质 (iii) (见(5.16))，当 $\eta \geqslant n$ 时，

$$\left(\frac{\eta}{n}\right)^{\frac{1}{n}} \Big/ g(\eta) \leqslant 1,$$

即

$$\frac{\eta}{g(\eta)} \leqslant n^{\frac{1}{n}}\eta^{1-\frac{1}{n}} \quad (\eta \geqslant n);$$

因此，由定理 5.2 可得

推论 5.5. 在定理 5.2 的假设条件下，只要

$$\nu(B-A) \leqslant \frac{\Delta_\nu(A)}{n}, \tag{5.24}$$

则有

$$s_A(B) \leqslant n^{\frac{1}{n}}(\Delta_\nu(A))^{1-\frac{1}{n}}[\nu(B-A)]^{\frac{1}{n}}. \tag{5.25}$$

5.3 正规性偏离度的估计

定理 5.3[102]. 设 $A \in \mathbb{C}^{n\times n}$, 则

$$\Delta_F(A) \leqslant \left(\frac{n^3-n}{12}\right)^{\frac{1}{4}} \|A^HA - AA^H\|_F^{\frac{1}{2}}. \tag{5.26}$$

证明:

设 $T = \Lambda + M$ 是 A 的 Schur 上三角形式,即 $U^HAU = T$, 其中 U 是酉阵, Λ 是对角阵, $M = (\mu_{ij})$ 是严格上三角阵. 令

$$\Gamma = T^HT - TT^H = (\gamma_{ij}).$$

1) 首先利用归纳法证明

$$\|M\|_F^2 \leqslant \gamma_{22} + 2\gamma_{33} + \cdots + (n-1)\gamma_{nn}, \tag{5.27}$$

其中

$$\gamma_{ii} = \sum_{k<i} |\mu_{ki}|^2 - \sum_{k>i} |\mu_{ik}|^2, i = 1, \cdots, n. \tag{5.28}$$

容易验证,当 $n = 2$ 时,有 $\|M\|_2^2 = \gamma_{22}$, 即(5.27)式成立. 现假定不等式 (5.27) 对于 n 阶矩阵已成立,考虑 $n+1$ 阶矩阵的情形如下.

把这个 $n+1$ 阶的 Schur 上三角形式记作 $\hat{T} = \hat{\Lambda} + \hat{M}$, 其中

$$\hat{M} = \begin{pmatrix} M & \begin{matrix}\mu_{1,n+1} \\ \vdots \\ \mu_{n,n+1}\end{matrix} \\ 0 \cdots 0 & 0 \end{pmatrix};$$

相应的 $\hat{\Gamma} = (\hat{\gamma}_{ij}) \in \mathbb{C}^{(n+1)\times(n+1)}$. 计算可知

$$\hat{\gamma}_{ii} = \gamma_{ii} - |\mu_{i,n+1}|^2, \quad i = 1, \cdots, n. \tag{5.29}$$

于是有

$$\|\hat{M}\|_F^2 = \|M\|_F^2 + |\mu_{1,n+1}|^2 + \cdots + |\mu_{n,n+1}|^2$$

$$\leqslant \gamma_{22} + 2\gamma_{33} + \cdots + (n-1)\gamma_{nn} + \hat{\gamma}_{n+1,n+1}$$

（据归纳法假设）

$$= \hat{\gamma}_{22} + |\mu_{2,n+1}|^2 + 2\hat{\gamma}_{33} + 2|\mu_{3,n+1}|^2 + \cdots$$

$$+ (n-1)\hat{\gamma}_{nn} + (n-1)|\mu_{n,n+1}|^2 + \hat{\gamma}_{n+1,n+1}$$

（据 (5.29)）

$$\leqslant \hat{\gamma}_{22} + 2\hat{\gamma}_{33} + \cdots + (n-1)\hat{\gamma}_{nn} + n\hat{\gamma}_{n+1,n+1}$$

$$(\because \hat{\gamma}_{n+1,n+1} = |\mu_{1,n+1}|^2 + |\mu_{2,n+1}|^2 + \cdots + |\mu_{n,n+1}|^2).$$

2) 从(5.27)式两端分别减去恒等式

$$0 = \gamma_{11} + \gamma_{22} + \cdots + \gamma_{nn}$$

的 $\frac{n-1}{2}$ 倍，得到

$$\|M\|_F^2 \leqslant \frac{1-n}{2}\gamma_{11} + \frac{3-n}{2}\gamma_{22} + \cdots + \frac{n-1}{2}\gamma_{nn}.$$

利用 Cauchy 不等式，有

$$\|M\|_F^4 \leqslant \left[\left(\frac{1-n}{2}\right)^2 + \left(\frac{3-n}{2}\right)^2 + \cdots + \left(\frac{n-1}{2}\right)^2\right](\gamma_{11}^2 + \cdots + \gamma_{nn}^2),$$

其中

$$\left(\frac{1-n}{2}\right)^2 + \left(\frac{3-n}{2}\right)^2 + \cdots + \left(\frac{n-1}{2}\right)^2 = \frac{n^3-n}{12},$$

$$\gamma_{11}^2 + \cdots + \gamma_{nn}^2 \leqslant \|\Gamma\|_F^2 = \|U^H(A^HA - AA^H)U\|_F^2 = \|A^HA - AA^H\|_F^2;$$

因此估计式(5.26) 成立. □

注 5.2. 如果在推论 5.5 中取范数 ν 为$\|\ \|_F$,并把估计式(5.26)用到(5.24)与(5.25)之中，便可得到 $s_A(B)$ 的一个可以计算的上界,这是比 Bauer-Fike 定理以及推广的 Bauer-Fike 定理(定理 5.1 及推论 5.2)优越之处;因为一般说来,算出 $\|Q\|_2\|Q^{-1}\|_2$ 的比较精

确的上界是困难的.

定理 5.4. 设 $A\in \mathbf{C}^{n\times n}$. 则

$$\Delta_F(A)\geqslant\left(\|A\|_F^2-\sqrt{\|A\|_F^4-\frac{1}{2}\|AA^H-A^HA\|_F^2}\right)^{\frac{1}{2}}. \tag{5.30}$$

证明:

首先指出一个恒等式,即

$$\|AA^H-A^HA\|_F^2=2(\|AA^H\|_F^2-\|A^2\|_F^2). \tag{5.31}$$

无妨设 A 已是 Schur 上三角形式, 即 $A=(\alpha_{ij})$, 当 $i>j$ 时, $\alpha_{ij}=0$. A 的严格上三角阵记作 M, 即

$$A=\mathrm{diag}(\alpha_{11},\cdots,\alpha_{nn})+M.$$

令

$$H=AA^H=(h_{ij}),\ G=A^2=(g_{ij}),$$

于是 $\left(以下 \sum\limits_i 表示 \sum\limits_{i=1}^n, \sum\limits_{i\leqslant j} 表示 \sum\limits_{1\leqslant i\leqslant j\leqslant n}, \sum\limits_i\left(\sum\limits_{i\leqslant j}\cdots\right) 表示 \sum\limits_{1\leqslant i\leqslant n}\left(\sum\limits_{i\leqslant j\leqslant n}\cdots\right), 等等\right)$

$$\|AA^H\|_F^2-\|A^2\|_F^2=\|H\|_F^2-\|G\|_F^2=\sum_{i,j}|h_{ij}|^2-\sum_{i\leqslant j}|g_{ij}|^2$$

$$\leqslant\sum_i|h_{ii}|^2+2\sum_{i<j}|h_{ij}|^2-\sum_i|g_{ii}|^2. \tag{5.32}$$

其中

$$\sum_i|h_{ii}|^2=\sum_i\left(\sum_{i\leqslant j}|\alpha_{ij}|^2\right)^2=\sum_i|\alpha_{ii}|^4+\sum_{i<j}|\alpha_{ij}|^4$$

$$+2\sum_{i\leqslant j<k}|\alpha_{ij}\alpha_{ik}|^2;$$

$$\sum_{i<j}|h_{ij}|^2=\sum_{i<j}\left|\sum_{j\leqslant k}\alpha_{ik}\bar{\alpha}_{jk}\right|^2$$

$$\leqslant\sum_{i<j}\left(\sum_{j\leqslant k}|\alpha_{ik}\alpha_{jk}|^2+2\sum_{j\leqslant k<l}|\alpha_{ik}\alpha_{jk}\alpha_{il}\alpha_{jl}|\right)$$

$$\leqslant \sum_{i<j\leqslant k} |\alpha_{ik}\alpha_{jk}|^2 + \sum_{i<j\leqslant k<l} |\alpha_{ik}\alpha_{jl}|^2$$

$$+ \sum_{i<j\leqslant k<l} |\alpha_{jk}\alpha_{il}|^2;$$

$$\sum_i |g_{ii}|^2 = \sum_i |\alpha_{ii}|^4;$$

代入(5.32)，得到

$$\|AA^H\|_F^2 - \|A^2\|_F^2 \leqslant \sum_{i<j} |\alpha_{ij}|^4 + 2 \sum_{i\leqslant j<k} |\alpha_{ij}\alpha_{jk}|^2$$

$$+ 2 \sum_{i<j\leqslant k} |\alpha_{ik}\alpha_{jk}|^2 + 2 \sum_{i<j\leqslant k<l} |\alpha_{ik}\alpha_{jl}|^2$$

$$+ 2 \sum_{i<j\leqslant k<l} |\alpha_{jk}\alpha_{il}|^2.$$

再注意到(适当调整指标符号)

$$\sum_{i\leqslant j<k} |\alpha_{ij}\alpha_{jk}|^2 = \sum_{i<j} |\alpha_{ii}\alpha_{ij}|^2 + \sum_{i=k<j<l} |\alpha_{ij}\alpha_{kl}|^2,$$

$$\sum_{i<j\leqslant k} |\alpha_{ik}\alpha_{jk}|^2 = \sum_{i\quad j} |\alpha_{ij}\alpha_{jj}|^2 + \sum_{i<k<j=l} |\alpha_{ij}\alpha_{kl}|^2,$$

$$\sum_{i<j\leqslant k<l} |\alpha_{ik}\alpha_{jl}|^2 = \sum_{i<k\leqslant j<l} |\alpha_{ij}\alpha_{kl}|^2,$$

$$\sum_{i<j\leqslant k<l} |\alpha_{jk}\alpha_{il}|^2 = \sum_{i<k<j} |\alpha_{kk}\alpha_{ij}|^2 + \sum_{i<k<l<j} |\alpha_{kl}\alpha_{ij}|^2,$$

所以

$$\|AA^H\|_F^2 - \|A^2\|_F^2 \leqslant \sum_{i<j} |\alpha_{ij}|^4$$

$$+ 2\left\{\sum_{\substack{i\leqslant j\\ k=i}} + \sum_{\substack{i\leqslant j\\ k=j}} + \sum_{\substack{i<j\\ i<k<j}}\right\} |\alpha_{ij}\alpha_{kk}|^2 + 2\left\{\sum_{i=k<j<l}\right.$$

$$\left. + \sum_{i<k<j=l} + \sum_{i<k\leqslant j<l} + \sum_{i\quad k<l<j}\right\} |\alpha_{ij}\alpha_{kl}|^2$$

$$\leqslant \sum_{i<j} |\alpha_{ij}|^4 + 2 \sum_{i<j} |\alpha_{ij}|^2 \sum_k |\alpha_{kk}|^2 + 2 \sum_{\substack{i<j\\ k\leqslant l\\ i\leqslant k}} |\alpha_{ij}\alpha_{kl}|^2. \tag{5.33}$$

另一方面，

$$2\|M\|_F^2\|A\|_F^2 = 2\sum_{i<j}|\alpha_{ij}|^2\cdot\left(\sum_{i<j}|\alpha_{ij}|^2+\sum_i|\alpha_{ii}|^2\right)$$

$$= 2\left(\sum_{i<j}|\alpha_{ij}|^2\right)^2 + 2\sum_{i<j}|\alpha_{ij}|^2\cdot\sum_k|\alpha_{kk}|^2$$

$$= 2\sum_{i<j}|\alpha_{ij}|^4 + 4\sum_{\substack{i<j\\k<l\\i\leqslant k}}|\alpha_{ij}\alpha_{kl}|^2 + 2\sum_{i<j}|\alpha_{ij}|^2\cdot\sum_k|\alpha_{kk}|^2. \tag{5.34}$$

由(5.33)和(5.34)得知

$$2\|M\|_F^2\|A\|_F^2 - (\|AA^H\|_F^2 - \|A^2\|_F^2)$$

$$\geqslant \sum_{i<j}|\alpha_{ij}|^4 + 2\sum_{\substack{i<j\\k<l\\i\leqslant k}}|\alpha_{ij}\alpha_{kl}|^2 = \left(\sum_{i<j}|\alpha_{ij}|^2\right)^2 = \|M\|_F^4.$$

将 (5.31) 代入上式，并令 $\|M\|_F^2=\xi$，$\|A\|_F^2=\alpha$ 和 $\|AA^H-A^HA\|_F^2=\beta\geqslant 0$，则得

$$\xi^2 - 2\alpha\xi + \frac{\beta}{2}\leqslant 0. \tag{5.35}$$

利用恒等式(5.31),可知(5.35)左端二项式的判别式为

$$2(2\|A\|_F^4 - \|AA^H - A^HA\|_F^2) \geqslant 4\|A^2\|_F^2\geqslant 0,$$

所以方程 $\xi^2-2\alpha\xi+\frac{\beta}{2}=0$ 有二实根 $\alpha\mp\sqrt{\alpha^2-\frac{\beta}{2}}$，从而使(5.35)成立的范围是

$$\alpha-\sqrt{\alpha^2-\frac{\beta}{2}}\leqslant\xi\leqslant\alpha+\sqrt{\alpha^2-\frac{\beta}{2}}.$$

由此立即导出 (5.30). □

应用定理 5.4 和推论 5.3，便可得到任一矩阵的特征值模的平方和的一个上界，即

推论 5.6. 设 $A\in C^{n\times n}$，$\lambda(A)=\{\lambda_i\}$. 则

$$\sum_{i=1}^n|\lambda_i|^2\leqslant\sqrt{\|A\|_F^4-\frac{1}{2}\|AA^H-A^HA\|_F^2}. \tag{5.36}$$

推论 5.6 改善了熟知的结果 $\sum_{i=1}^{n}|\lambda_i|^2 \leqslant \|A\|_F^2$. 定理 5.4 的上述证明是作者[12]给出的.

5.4 Henrici 定理(续)

七十年代末以来,Bhatia 等人进一步发展了 Ostrowski 和 Henrici 的结果，得到了矩阵特征值扰动的可以计算的一些上界. Elsner [86] 指出，Bhatia 等人的结果，可以从 Henrici 定理导出,下面就来讨论这个问题.

Henrici 定理(定理 5.2)给出了任一 n 阶矩阵 B 对于 A 的谱改变量的上界,即

$$\left.\begin{aligned} &s_A(B) \leqslant \frac{\Delta_\nu(A)}{g\left(\dfrac{\Delta_\nu(A)}{\nu(B-A)}\right)} \equiv s_n(\Delta, r), \\ &\Delta = \Delta_\nu(A), \quad r = \nu(B-A), \end{aligned}\right\} \tag{5.37}$$

其中 ν 是矩阵谱范数的任一种强范数，$\Delta_\nu(A)$ 是 A 对于范数 ν 的正规性偏离度，$g(\eta)$ 是 $g + g^2 + \cdots + g^n = \eta(\eta > 0)$ 的唯一正根.

以下假定 $\Delta, r > 0$, 即 A 非正规,并且 $B \neq A$.

令 $\lambda_M(A)$ 表示矩阵 A 的最大实特征值. 记

$$L(\Delta, r) = \begin{pmatrix} 0 & \Delta & & \\ & \ddots & \ddots & \\ & & 0 & \Delta \\ r & \cdots & r & r \end{pmatrix} \in \mathbb{R}^{n\times n}, \quad \Delta, r > 0. \tag{5.38}$$

首先证明一条引理.

引理 5.2. (5.37)式所示的

$$s_n(\Delta, r) \equiv \frac{\Delta}{g\left(\dfrac{\Delta}{r}\right)}$$

有下列性质:

(i) $s_n(\Delta, r)$ 可以表示为

$$s_n(\Delta, r) = \lambda_M(L(\Delta, r)), \tag{5.39}$$

其中 $L(\Delta,r)$ 如(5.38)式所示;

(ii) $s_n(\Delta,r)$ 对Δ和 r 单调递增;

(iii) $s_n(\Delta, r)/r^{\frac{1}{n}}$ 对Δ和 r 单调递增。

证明:

(i) 因为 $L(\Delta,r)$ 的特征多项式为

$$p(\lambda) \equiv \det(\lambda I - L(\Delta,r))$$

$$= \Delta^n \det\left(\frac{\lambda}{\Delta} I - \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ \frac{r}{\Delta} & \cdots & \frac{r}{\Delta} & \frac{r}{\Delta} \end{pmatrix}\right)$$

$$= \Delta^n \left[\left(\frac{\lambda}{\Delta}\right)^n - \frac{r}{\Delta}\left(\frac{\lambda}{\Delta}\right)^{n-1} - \cdots - \frac{r}{\Delta}\left(\frac{\lambda}{\Delta}\right) - \frac{r}{\Delta}\right]^*.$$

所以若令

$$\frac{\lambda}{\Delta} = \lambda', \quad \frac{\Delta}{r} = \alpha,$$

则 $L(\Delta,r)$ 的特征方程为

$$\lambda'^n - \frac{1}{\alpha}\lambda'^{n-1} - \cdots - \frac{1}{\alpha}\lambda' - \frac{1}{\alpha} = 0, \ \lambda = \Delta\lambda'. \qquad (5.40)$$

显然(5.40)的根均不为零,因此如果记 $\lambda' = \frac{1}{g}$, 则 g 应满足

$$\frac{1}{g^n}\left[1 - \frac{1}{\alpha}(g + g^2 + \cdots + g^n)\right] = 0,$$

或者等价地, g 应满足

$$g + g^2 + \cdots + g^n = \alpha, \ \alpha > 0; \qquad (5.41)$$

---

* 设 $f(\lambda) = \lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_n$. 令

$$L(f) = \begin{pmatrix} 0 & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ -\alpha_n & \cdots & & -\alpha_2 & -\alpha_1 \end{pmatrix}.$$

$L(f)$ 叫做多项式 f 的友阵. $L(f)$ 的特征多项式即 $f(\lambda)$.

并且 $L(\Delta,r)$ 的最大实特征值，就是方程 (5.41) 的最大根(唯一正根)的倒数的Δ倍. 因为方程 (5.41) 的这个唯一正根(即最大根)记作 $g(\alpha)$，所以

$$\lambda_M(L(\Delta,r))=\frac{\Delta}{g(\alpha)}=\frac{\Delta}{g\left(\frac{\Delta}{r}\right)}=s_n(\Delta,r).$$

(ii) 根据 $g(\eta)$ 与 $\frac{\eta}{g(\eta)}$ 的单调递增性可知，如果 $\Delta\leqslant\Delta'$，则

$$s_n(\Delta,r)=\frac{\frac{\Delta}{r}}{g\left(\frac{\Delta}{r}\right)}\cdot r\leqslant\frac{\frac{\Delta'}{r}}{g\left(\frac{\Delta'}{r}\right)}\cdot r=s_n(\Delta',r);$$

如果 $r\leqslant r'$，则

$$s_n(\Delta,r)=\frac{\Delta}{g\left(\frac{\Delta}{r}\right)}\leqslant\frac{\Delta}{g\left(\frac{\Delta}{r'}\right)}=s_n(\Delta,r').$$

(iii) 根据 $\frac{\eta}{g(\eta)}$ 的单调递增性与 $\eta^{\frac{1}{n}}\big/g(\eta)$ 的单调递降性可知，如果 $\Delta\leqslant\Delta'$，则

$$\frac{s_n(\Delta,r)}{r^{\frac{1}{n}}}=\frac{\Delta}{r^{\frac{1}{n}}g\left(\frac{\Delta}{r}\right)}=r^{1-\frac{1}{n}}\cdot\frac{\frac{\Delta}{r}}{g\left(\frac{\Delta}{r}\right)}$$

$$\leqslant r^{1-\frac{1}{n}}\cdot\frac{\frac{\Delta'}{r}}{g\left(\frac{\Delta'}{r}\right)}=\frac{s_n(\Delta',r)}{r^{\frac{1}{n}}};$$

如果 $r\leqslant r'$，则

$$\frac{s_n(\Delta,r)}{r^{\frac{1}{n}}}=\frac{\left(\frac{\Delta}{r}\right)^{\frac{1}{n}}}{g\left(\frac{\Delta}{r}\right)}\cdot\Delta^{1-\frac{1}{n}}\leqslant\frac{\left(\frac{\Delta}{r'}\right)^{\frac{1}{n}}}{g\left(\frac{\Delta}{r'}\right)}\cdot\Delta^{1-\frac{1}{n}}$$

$$= \frac{s_n(\Delta, r')}{r'^{\frac{1}{n}}}. \quad \square$$

定理 5.5 (Bhatia,Friedland, [59]). 设 $A, B \in \mathbf{C}^{n\times n}$. 则

$$s_A(B) \leqslant n^{\frac{1}{n}}(2m_2)^{1-\frac{1}{n}} \|B - A\|_2^{\frac{1}{n}}, \tag{5.42}$$

其中

$$m_2 = \max\{\|A\|_2, \|B\|_2\}. \tag{5.43}$$

证明:

首先指出两点:

(i) $s_A(B) \leqslant \max\limits_{\substack{\lambda_i \in \lambda(A)\\ \mu_j \in \lambda(B)}} |\lambda_i - \mu_j| \leqslant \rho(A) + \rho(B)$

$$\leqslant \|A\|_2 + \|B\|_2 \leqslant 2m_2;$$

(ii) 据 A 的 Schur 分解 $U^H AU = \Lambda + M$ 可知

$$\Delta_2(A) \leqslant \|M\|_2 \leqslant \|A\|_2 + \|\Lambda\|_2 \leqslant 2m_2.$$

以下分两种情形证明定理之结论:

1) 当 $\|B - A\|_2 \geqslant \frac{2m_2}{n}$ 时,有

$$(2m_2)^{\frac{1}{n}} \leqslant n^{\frac{1}{n}}\|B - A\|_2^{\frac{1}{n}},$$

从而

$$s_A(B) \leqslant 2m_2 \leqslant (2m_2)^{1-\frac{1}{n}} \cdot n^{\frac{1}{n}}\|B - A\|_2^{\frac{1}{n}};$$

2) 当 $\|B - A\|_2 < \frac{2m_2}{n}$ 时,记 $\Delta = \Delta_2(A)$ 和 $r = \|B-A\|_2$,

有

$$s_A(B) \leqslant s_n(\Delta, r) \leqslant \frac{s_n(2m_2, r)}{r^{\frac{1}{n}}} \cdot r^{\frac{1}{n}}$$

(据 $s_n(\Delta, r)$ 对 Δ 的单调性)

$$\leqslant \frac{s_n\left(2m_2, \frac{2m_2}{n}\right)}{\left(\frac{2m_2}{n}\right)^{\frac{1}{n}}} \cdot r^{\frac{1}{n}} \left(据 \frac{s_n(\Delta, r)}{r^{\frac{1}{n}}} 对 r 的单调性\right)$$

$$= n^{\frac{1}{n}}(2m_2)^{-\frac{1}{n}}\lambda_M\left(\begin{pmatrix} 0 & 2m_2 & & \\ & 0 & \ddots & \\ & & \ddots & \ddots \\ & & 0 & 2m_2 \\ \frac{2m_2}{n} & \cdots & \frac{2m_2}{n} & \frac{2m_2}{n} \end{pmatrix}\right)\cdot r^{\frac{1}{n}}$$

$$= n^{\frac{1}{n}}(2m_2)^{1-\frac{1}{n}}\ r^{\frac{1}{n}}\lambda_M\left(\begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ \frac{1}{n} & \cdots & \frac{1}{n} & \frac{1}{n} \end{pmatrix}\right)$$

$$= n^{\frac{1}{n}}(2m_2)^{1-\frac{1}{n}}\ \|B - A\|_2^{\frac{1}{n}}.$$

上面最后一个等式是根据 $L\left(1,\frac{1}{n}\right)$的特征方程

$$\lambda^n - \frac{1}{n}\lambda^{n-1} - \cdots - \frac{1}{n}\lambda - \frac{1}{n} = 0$$

有唯一的非负根 $\lambda = 1$. □

定理 5.6 (Bhatia, Mukherjea, [60], Elsner, [86]). 设 A, $B \in \mathbf{C}^{n\times n}$, 则

$$s_A(B) \leqslant n^{\frac{1}{2n}}\left[\left(1+\frac{1}{\sqrt{n}}\right)^n - 1\right]^{\frac{1}{n}} m_F^{1-\frac{1}{n}}\|B - A\|_F^{\frac{1}{n}}, \tag{5.44}$$

其中

$$m_F = \max\{\|A\|_F, \|B\|_F\}. \tag{5.45}$$

证明:

首先指出两点:

(i) 如果把 A 与 B 的特征值分别排列为

$$|\lambda_1| \leqslant |\lambda_2| \leqslant \cdots \leqslant |\lambda_n|, |\mu_1| \geqslant |\mu_2| \geqslant \cdots \geqslant |\mu_n|,$$

利用 $\sum_{i=1}^{n} |\lambda_i|^2 \leqslant m_F^2$ 和 $\sum_{i=1}^{n} |\mu_i|^2 \leqslant m_F^2$, 通过归纳法可证

$$|\lambda_i - \mu_i| \leqslant \left(1 + \frac{1}{\sqrt{n}}\right) m_F,$$

因而

$$s_A(B) \leqslant \left(1+\frac{1}{\sqrt{n}}\right) m_F;$$

(ii) 显然有

$$\Delta_F(A) \leqslant m_F.$$

以下分两种情形证明定理之结论:

1) 当 $\|B-A\|_F \geqslant \tau^{-n}\left(1+\frac{1}{\sqrt{n}}\right)^n m_F$ 时(其中 τ 待定),有

$$s_A(B) \leqslant \left(1+\frac{1}{\sqrt{n}}\right) m_F \leqslant \tau m_F^{1-\frac{1}{n}} \|B-A\|_F^{\frac{1}{n}}; \quad (5.46)$$

2) 当 $\|B-A\|_F < \tau^{-n}\left(1+\frac{1}{\sqrt{n}}\right)^n m_F$ 时,记 $\Delta=\Delta_F(A)$ 和 $r=\|B-A\|_F$,于是有

$$s_A(B) \leqslant s_n(\Delta, r) \leqslant \frac{s_n\left(m_F, \tau^{-n}\left(1+\frac{1}{\sqrt{n}}\right)^n m_F\right)}{\tau^{-1}\left(1+\frac{1}{\sqrt{n}}\right) m_F^{\frac{1}{n}}} \cdot r^{\frac{1}{n}}$$

$$\left(\text{据} \frac{s_n(\Delta, r)}{r^{\frac{1}{n}}} \text{ 对 } r \text{ 的单调性}\right)$$

$$\leqslant m_F^{1-\frac{1}{n}} \tau \cdot r^{\frac{1}{n}} \cdot \frac{s_n\left(1, \tau^{-n}\left(1+\frac{1}{\sqrt{n}}\right)^n\right)}{1+\frac{1}{\sqrt{n}}}. \quad (5.47)$$

现取 τ 使得

$$\frac{s_n\left(1, \tau^{-n}\left(1+\frac{1}{\sqrt{n}}\right)^n\right)}{1+\frac{1}{\sqrt{n}}} = 1$$

即

$$\left(1+\frac{1}{\sqrt{n}}\right) g\left(\frac{\tau^n}{\left(1+\frac{1}{\sqrt{n}}\right)^n}\right) = 1,$$

也就是取 τ 使得

$$g + g^2 + \cdots + g^n = \frac{\tau^n}{\left(1 + \frac{1}{\sqrt{n}}\right)^n} \tag{5.48}$$

的非负根 g 满足 $g = \frac{1}{1 + \frac{1}{\sqrt{n}}}$. 代入(5.48)可算出

$$\begin{aligned} \tau^n &= \left(1 + \frac{1}{\sqrt{n}}\right)^n \left[\left(1 + \frac{1}{\sqrt{n}}\right)^{-1} + \cdots + \left(1 + \frac{1}{\sqrt{n}}\right)^{-n}\right] \\ &= \sum_{k=0}^{n-1}\left(1 + \frac{1}{\sqrt{n}}\right)^k = \left[\left(1 + \frac{1}{\sqrt{n}}\right)^n - 1\right]\sqrt{n}, \end{aligned}$$

即

$$\tau = n^{\frac{1}{2n}}\left[\left(1 + \frac{1}{\sqrt{n}}\right)^n - 1\right]^{\frac{1}{n}}.$$

将上式所示的 τ 分别代入(5.46)和(5.47),便得出(5.44). □

注 5.3. 容易验证

$$n^{\frac{1}{2}}\left[\left(1 + \frac{1}{\sqrt{n}}\right)^n - 1\right] \leqslant \sum_{k=1}^{n} k^{1-\frac{k}{2}}\binom{n}{k},$$

所以从(5.43)可以推导出 Bhatia 和 Mukherjea 原来的估计式

$$s_A(B) \leqslant \left[\sum_{k=1}^{n} k^{1-\frac{k}{2}}\binom{n}{k}\right]^{\frac{1}{n}} m_F^{1-\frac{1}{n}}\|B - A\|_F^{\frac{1}{n}}. \tag{5.49}$$

注 5.4. 应用定理 1.2, 从定理 5.5 和定理 5.6 得到估计式

$$v(A, B) \leqslant (2n - 1)n^{\frac{1}{n}}(2m_2)^{1-\frac{1}{n}}\|B - A\|_2^{\frac{1}{n}} \tag{5.50}$$

和

$$v(A, B) \leqslant (2n - 1)n^{\frac{1}{2n}}\left[\left(1 + \frac{1}{\sqrt{n}}\right)^n - 1\right]^{\frac{1}{n}} m_F^{1-\frac{1}{n}}\|B - A\|_F^{\frac{1}{n}}, \tag{5.51}$$

其中 m_2 与 m_F 分别如(5.43)与(5.45)所示.

5.5 举例

下面的两个例子，引自 Henrici (1962) 的论文[102]。引用它们，在于说明如何应用 Ostrowski、Henrici 和 Bhatia 等人的结果（定理 1.1，定理 5.2，定理 5.5 和定理 5.6）去估算谱改变量 $s_A(B)$ 的上界。

例 5.1. $A=\begin{pmatrix}1 & 1\\ 0 & 1-10^{-4}\end{pmatrix}$, $B=\begin{pmatrix}1 & 1\\ -10^{-2} & 1-10^{-4}\end{pmatrix}$.

计算可知:

$$\lambda(A)=\{1, 0.9999\},$$

$$\lambda(B)=\{0.99995\pm 0.0999998i\},$$

$$s_A(B)=0.099999812,$$

$$m_\alpha=1,\ m_2=1.6207621,\ m_F=1.7320219,$$

$$\|B-A\|_\alpha=0.005,\ \|B-A\|_2=\|B-A\|_F=10^{-2},$$

$$\Delta_2(A)=\Delta_F(A)=1,\ \eta=\frac{\Delta_2(A)}{\|B-A\|_2}=\frac{\Delta_F(A)}{\|B-A\|_F}=10^2,$$

$g+g^2=\eta$ 的非负解 $g(\eta)=9.5124922$.

定理 5.1 (Ostrowski):

$$s_A(B)\leqslant 4\times 0.005^{\frac{1}{2}}=0.2828427;$$

定理 5.2 (Henrici):

$$s_A(B)\leqslant\frac{\eta}{g(\eta)}\cdot\|B-A\|_2=\frac{\eta}{g(\eta)}\cdot\|B-A\|_F$$

$$=\frac{10^2}{9.5124922}\times 10^{-2}=0.1051249;$$

定理 5.5 (Bhatia-Friedland):

$$s_A(B)\leqslant 2^{\frac{1}{2}}\times(2\times 1.6207621)^{\frac{1}{2}}\times(10^{-2})^{\frac{1}{2}}=0.2546183;$$

定理 5.6 (Bhatia-Mukherjea-Elsner):

$$s_A(B)\leqslant 2^{\frac{1}{4}}\times\left[\left(1+\frac{1}{\sqrt{2}}\right)^2-1\right]^{\frac{1}{2}}\times 1.7320219^{\frac{1}{2}}$$

$$\times(10^{-2})^{\frac{1}{2}}=0.1645328.$$

例 5.2. $A=\begin{pmatrix}1 & 10^{-4}\\ 0 & 1\end{pmatrix}$, $B=\begin{pmatrix}1+2^{-\frac{1}{2}}10^{-2} & 10^{-4}\\ 0 & 1-2^{-\frac{1}{2}}10^{-2}\end{pmatrix}$.

显然有

$$\lambda(A)=\{1,1\},\ \lambda(B)=\{1.00707106, 0.992929\}$$

和

$$s_A(B)=0.00707106.$$

计算可知

$m_\alpha=1.00707106$, $m_2=1.0070711$, $m_F=1.4142489$,

$$\|B-A\|_\alpha=\|B-A\|_2=0.00707106,$$

$\|B-A\|_F=0.0099999$, $\Delta_2(A)=\Delta_F(A)=10^{-4}$,

$$\eta_2=\frac{\Delta_2(A)}{\|B-A\|_2}=0.0141421,\quad \eta_F=\frac{\Delta_F(A)}{\|B-A\|_F}=0.0100001,$$

$g+g^2=\eta_2$ 的非负解 $g(\eta_2)=0.0139475$,

$g+g^2=\eta_F$ 的非负解 $g(\eta_F)=0.00990204$.

定理 5.1 (Ostrowski):

$s_A(B)\leqslant 4\times 1.00707106^{\frac{1}{2}}\times 0.00707106^{\frac{1}{2}}=0.1193398$;

定理 5.2 (Henrici):

$$s_A(B)\leqslant \frac{\eta_2}{g(\eta_2)}\|B-A\|_2=\frac{0.0141421}{0.0139475}\times 0.00707106$$

$$=0.0071697;$$

$$s_A(B)\leqslant \frac{\eta_F}{g(\eta_F)}\|B-A\|_F=\frac{0.0100001}{0.00990204}\times 0.0099999$$

$$=0.0100989;$$

定理 5.5 (Bhatia-Friedland):

$$s_A(B)\leqslant 2^{\frac{1}{2}}\times(2\times 1.0070711)^{\frac{1}{2}}\times(0.00707106)^{\frac{1}{2}}$$

$$=0.168772;$$

定理 5.6 (Bhatia-Mukherjea-Elsner):

$$s_A(B)\leqslant 2^{\frac{1}{4}}\times\left[\left(1+\frac{1}{\sqrt{2}}\right)^2-1\right]^{\frac{1}{2}}\times 1.4142489^{\frac{1}{2}}$$

$$\times 0.0099999^{\frac{1}{2}}=0.1956651.$$

习题

1. 对于实变数 $\xi \geqslant 0$，定义函数 f 如 (5.13) 式所示。g 表示 f 的反函数，如(5.14)式所示。证明 f 与 g 有 §5 中所述的性质 (i)—(iv)。

2. 设 $A \in \mathbf{C}^{n\times n}$ 为非正规阵。$\boldsymbol{x} \in \mathbf{C}^n$，$\|\boldsymbol{x}\|_2 = 1$，$\lambda \in \mathbf{C}$，适合 $A\boldsymbol{x} \neq \lambda\boldsymbol{x}$。令

$$\eta = \frac{\Delta_2(A)}{\|A\boldsymbol{x} - \lambda\boldsymbol{x}\|_2}.$$

试证：存在 $\lambda_i \in \lambda(A)$，使得

$$|\lambda_i - \lambda| \leqslant \frac{\eta}{g(\eta)} \|A\boldsymbol{x} - \lambda\boldsymbol{x}\|_2,$$

其中 $g(\eta)$ 表示 $g + g^2 + \cdots + g^n = \eta(\eta > 0)$ 的唯一非负解。

3. 设 $A \in \mathbf{C}^{n\times n}$，$\lambda(A) = \{\lambda_i\}$。试证：对于 A 的值域 $F(A)$ 中的任一点 ξ，必有 $\lambda(A)$ 的凸包 $\mathscr{H}(\lambda_1, \cdots, \lambda_n)$ 中的点 η，使得

$$|\xi - \eta| \leqslant \sqrt{\frac{1 - n^{-1}}{2}}\Delta_F(A).$$

其中 $\Delta_F(A)$ 表示 A 对于范数 $\|\ \ \|_F$ 的正规性偏离度。

4. 设 $A \in \mathbf{C}^{n\times n}$，$\lambda(A) = \{\lambda_i\}$，$\sigma(A) = \{\sigma_i\}$。则存在 $\{\lambda_i\}$ 与 $\{\sigma_i\}$ 的适当排列，使得 $|\sigma_i - |\lambda_i|| \leqslant \Delta_2(A)$。

§6 条 件 数

6.1 特征值问题病态程度的数据标准

如果特征值问题的计算结果对于初始数据的一个微小扰动是很敏感的，则称该问题是病态的；否则称为良态的。本节讨论这种病态程度的数据标准，即特征值问题的条件数。

定理 3.6 和定理 4.7 说明，Hermite 阵的特征值是良态的；推论 4.2 说明，正规阵的特征值也是良态的。但是定理 4.5 表明，即

使是可对角化阵 $A = Q^{-1}\Lambda Q$（其中 $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$），如果 $\|Q\|_2\|Q^{-1}\|_2$ 很大，则 A 的微小扰动，可能导致它的特征值的很大改变。于是引进了 A 的谱条件数的概念。

设 $A \in \mathbf{C}^{n\times n}$，它的 Jordan 标准形为 J。令

$$\mathscr{Q}_A = \{Q \in \mathbf{C}^{n\times n}: Q^{-1}AQ = J\} \tag{6.1}$$

和

$$\nu(A) = \inf_{Q \in \mathscr{Q}_A} \|Q\|_2\|Q^{-1}\|_2, \nu_F(A) = \inf_{Q \in \mathscr{Q}_A} \|Q\|_F\|Q^{-1}\|_F. \tag{6.2}$$

上式中的 $\nu(A)$ 和 $\nu_F(A)$ 叫做 A 的谱条件数。

容易看出，$\nu_F(A) \geqslant \nu(A) \geqslant 1$，并且酉相似变换不改变 A 的谱条件数 $\nu(A)$ 和 $\nu_F(A)$。

推论 5.2 表明，对任一 $A \in \mathbf{C}^{n\times n}$，如果 A 的 Jordan 分解式为 $A = QJQ^{-1}$，J 为 A 的 Jordan 标准形，则 $s_A(B)$ 的上界中包含因子 $(\|Q\|_2\|Q^{-1}\|_2)^{\frac{1}{l}}$，其中 l 是 J 中最大 Jordan 块的阶数。因此 $\nu(A)$ 和 $\nu_F(A)$ 反映了 A 的特征值问题的病态程度。

另一方面，从 § 2 中 2.2 的讨论得知，如果 $A \in \mathbf{C}^{n\times n}$ 是一可正规化阵，其特征值为 $\lambda_1, \cdots, \lambda_n$，$\lambda_1$ 是它的 p 重特征值，并且 $B = (\beta_{ij}) \in \mathbf{C}^{n\times n}$，适合 $|\beta_{ij}| \leqslant 1\ (1 \leqslant i, j \leqslant n)$，则 $A + \varepsilon B$ $(0 < \varepsilon \ll 1)$ 必有 p 个特征值 $\mu_1(\varepsilon), \cdots, \mu_p(\varepsilon)$，满足

$$|\mu_i(\varepsilon) - \lambda_1| \lesssim \frac{np\varepsilon}{\min\limits_{1\leqslant i\leqslant p} s_i}, \quad i = 1, \cdots, p. \tag{6.3}$$

上式中的 $s_i (i = 1, 2, \cdots, p)$ 如下定义：

$$s_i = \frac{\boldsymbol{y}_i^H \boldsymbol{x}_i}{\|\boldsymbol{x}_i\|_2\|\boldsymbol{y}_i\|_2} > 0, \tag{6.4}$$

其中 $\boldsymbol{x}_i$ 与 $\boldsymbol{y}_i$ 是 A 对应于 λ_i 的适当选取的右特征向量与左特征向量（见(2.6)—(2.8)）。

(6.3)式表明，按照(6.4)式定义的一组正数 $\left\{\frac{1}{s_i}\right\}$，也反映了 A

的特征值问题的病态程度. 所以,$\frac{1}{s_1},\cdots,\frac{1}{s_n}$ 也叫做特征值问题 $Ax=\lambda x$ 的条件数.

从(6.4)可以看出,

$$s_i = \cos\theta(R(\boldsymbol{x}_i), R(\boldsymbol{y}_i)), \ i=1,2,\cdots,n.$$

因此,$\frac{1}{s_i}\geqslant 1$,并且酉相似变换不改变 $Ax=\lambda x$ 的条件数$\frac{1}{s_i}$, $i=1,\cdots,n$.

设 $\lambda\in\lambda(A)$. 任取 A 对应于 λ 的右特征向量 $\boldsymbol{x}$ 与左特征向量 $\boldsymbol{y}^H$. 如果令 $s=\boldsymbol{y}^H\boldsymbol{x}/\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2$, 则这样定义的 s 既不是唯一的,也不保证 $s\neq 0$. 但如果 λ 是 A 的单特征值,则如此定义的 s 必不为零,而且$\frac{1}{|s|}$唯一确定. 这一事实可叙述为

定理 6.1. 设 $A\in\mathbb{C}^{n\times n}$, λ 是 A 的单特征值, $\boldsymbol{x}$ 与 $\boldsymbol{y}^H$ 分别是 A 属于 λ 的右特征向量与左特征向量,即

$$A\boldsymbol{x}=\lambda\boldsymbol{x}, \quad \boldsymbol{y}^H A=\lambda\boldsymbol{y}^H.$$

令

$$s=\frac{\boldsymbol{y}^H\boldsymbol{x}}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2},$$

则 $s\neq 0$, 并且 $|s|$ 是唯一确定的.

证明:

容易证明, A 属于单特征值 λ 的右特征向量 $\boldsymbol{x}$ 和左特征向量 $\boldsymbol{y}$, 在相差一个复数因子的条件下,都是唯一确定的. 因此 $|s|$ 唯一确定. 以下证明 $\boldsymbol{y}^H\boldsymbol{x}\neq 0$.

不失一般性, 可以假定 $\|\boldsymbol{x}\|_2=1$. 构造酉阵 $X=(\boldsymbol{x},X_1)$, 于是有

$$X^H A X=\begin{pmatrix}\lambda & \boldsymbol{a}^H\\ 0 & C\end{pmatrix}.$$

令

$$\boldsymbol{y}^H=(1,\boldsymbol{a}^H(\lambda I-C)^{-1})X^H.$$

直接验证可知

$$\boldsymbol{y}^H A = \lambda \boldsymbol{y}^H, \quad \boldsymbol{y}^H \boldsymbol{x} \neq 0. \qquad \square$$

下述定理将进一步阐明上面所说的 $\frac{1}{|s|}$ 在特征值问题扰动理论中的意义.

定理 6.2. 设 $A, B \in \mathbb{C}^{n\times n}$, λ 是 A 的单特征值, $\boldsymbol{x}$ 与 $\boldsymbol{y}^H$ 分别是 A 属于 λ 的右特征向量与左特征向量. 则存在 $\varepsilon_0 > 0$ 以及在 $|\varepsilon| \leqslant \varepsilon_0$ 上的解析函数 $\lambda(\varepsilon)$, 满足

$$\lambda(0) = \lambda, \quad \lambda'(0) \equiv \left.\frac{d\lambda(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = \frac{\boldsymbol{y}^H B \boldsymbol{x}}{\boldsymbol{y}^H \boldsymbol{x}};$$

在 $|\varepsilon| \leqslant \varepsilon_0$ 上, $\lambda(\varepsilon)$ 是 $A + \varepsilon B$ 的单特征值;并且 $\lambda(\varepsilon)$ 可以表示为

$$\lambda(\varepsilon) = \lambda + \varepsilon \frac{\boldsymbol{y}^H B \boldsymbol{x}}{\boldsymbol{y}^H \boldsymbol{x}} + O(|\varepsilon|^2). \tag{6.5}$$

证明:

首先把 $A + \varepsilon B$ 的特征多项式记作

$$\varphi_\varepsilon(z) = \det[zI - (A + \varepsilon B)].$$

$\varphi_\varepsilon(z)$ 显然是 ε 和 z 的解析函数. 令

$$\mathscr{D}_r = \{z \in \mathbb{C}: |z - \lambda| \leqslant r\}.$$

取 r 足够小, 使得 A 在 $\mathscr{D}_r$ 上只有特征值 λ. $\mathscr{D}_r$ 的边界记为 $\partial\mathscr{D}_r$. 于是有

$$\min_{z \in \partial\mathscr{D}_r} |\varphi_0(z)| = \gamma > 0.$$

因为 $\varphi_\varepsilon(z)$ 对于 ε 连续, 所以存在 $\varepsilon_0 > 0$, 使得对于每一个适合 $|\varepsilon| \leqslant \varepsilon_0$ 的 ε, $\varphi_\varepsilon(z)$ 在 $\mathscr{D}_r$ 上只有一个零点,并且

$$\min_{\substack{z \in \partial\mathscr{D}_r \\ |\varepsilon| \leqslant \varepsilon_0}} |\varphi_\varepsilon(z)| > 0.$$

于是由留数理论可知*, $\varphi_\varepsilon(z)$ 在 $\mathscr{D}_r$ 内的零点 $\lambda(\varepsilon)$ 可表示成

---

* 函数论中的一条定理: 设 $\mathscr{D}$ 是 $\mathbb{C}$ 中的一个闭区域,其边界为 $\partial\mathscr{D}$. 如果函数 $f(z)$ 在 $\mathscr{D}$ 上解析,并且当 $z \in \partial\mathscr{D}$ 时, $f(z) \neq \alpha$, 则

$$\frac{1}{2\pi i} \oint_{\partial\mathscr{D}} z \frac{f'(z)}{f(z) - \alpha} dz = n_1 z_1 + \cdots + n_m z_m.$$

其中 $f'(z) \equiv df(z)/dz$, $z_1, \cdots, z_m$ 分别是 $f(z) - \alpha$ 在 $\mathscr{D}$ 内的 n_1 重零点,$\cdots$, n_m 重零点.

$$\lambda(\varepsilon) = \frac{1}{2\pi_i} \oint_{\partial\mathscr{D}_r} \frac{z\varphi'_\varepsilon(z)}{\varphi_\varepsilon(z)} dz,$$

其中 $\varphi'_\varepsilon(z) = d\varphi_\varepsilon(z)/dz$.

注意到 $\frac{z\varphi'_\varepsilon(z)}{\varphi_\varepsilon(z)}$ 和 $\frac{d}{d\varepsilon}\left(\frac{z\varphi'_\varepsilon(z)}{\varphi_\varepsilon(z)}\right)$ 在 $\partial\mathscr{D}_r$ 上连续，从而可以利用微分与积分次序交换定理得知，$\lambda(\varepsilon)$ 是 $|\varepsilon| \leqslant \varepsilon_0$ 上的解析函数．所以 $\lambda(\varepsilon)$ 可以表示成

$$\lambda(\varepsilon) = \lambda(0) + \lambda'(0)\varepsilon + O(|\varepsilon|^2),\ \lambda(0) = \lambda,\ |\varepsilon| \leqslant \varepsilon_0. \tag{6.6}$$

设 $\boldsymbol{x}(\varepsilon)$ 与 $\boldsymbol{y}(\varepsilon)^H$ 分别是 $A + \varepsilon B$ 属于特征值 $\lambda(\varepsilon)$ 的右特征向量与左特征向量．$\boldsymbol{x}(\varepsilon)$ 与 $\boldsymbol{y}(\varepsilon)$ 可以如下确定：据题设，$\mathrm{rank}(\lambda I - A) = n - 1$．因此 $\lambda I - A$ 必有一个 $n-1$ 阶非奇异子阵．假定这个子阵位于 $\lambda I - A$ 的第 2 至第 n 行．用 $\xi_1, \cdots, \xi_n$ 表示 $\lambda I - A$ 的第一行元素的代数余子式，则有 $(\xi_1, \cdots, \xi_n) \neq 0$ 和 $(\lambda I - A)(\xi_1, \cdots, \xi_n)^T = 0$．已知 λ 是 A 的单特征值，相应的右特征向量 $\boldsymbol{x}$ 在容许相差一个复数因子的条件下是唯一确定的，所以无妨假设 $\boldsymbol{x} = (\xi_1, \cdots, \xi_n)^T$．根据特征值与行列式对于矩阵元素的连续性，可取 ε_0 足够小，使得当 $|\varepsilon| \leqslant \varepsilon_0$ 时，$\lambda(\varepsilon)I - (A + \varepsilon B)$ 的第 2 行至第 n 行线性无关；分别把 $\lambda(\varepsilon)I - (A + \varepsilon B)$ 的第一行诸元素的代数余子式记为 $\xi_1(\varepsilon), \cdots, \xi_n(\varepsilon)$，并令 $\boldsymbol{x}(\varepsilon) = (\xi_1(\varepsilon), \cdots, \xi_n(\varepsilon))^T$，则显然有 $\boldsymbol{x}(\varepsilon) \neq 0$，并且

$$(A + \varepsilon B)\boldsymbol{x}(\varepsilon) = \lambda(\varepsilon)\boldsymbol{x}(\varepsilon), \tag{6.7}$$

即 $\boldsymbol{x}(\varepsilon)$ 是 $A + \varepsilon B$ 属于 $\lambda(\varepsilon)$ 的右特征向量．同理可定出 $A + \varepsilon B$ 属于 $\lambda(\varepsilon)$ 的左特征向量 $\boldsymbol{y}(\varepsilon)^H$:

$$\boldsymbol{y}(\varepsilon)^H(A + \varepsilon B) = \lambda(\varepsilon)\boldsymbol{y}(\varepsilon)^H.$$

容易看出，上面所得到的 $\boldsymbol{x}(\varepsilon)$ 与 $\boldsymbol{y}(\varepsilon)$，它们的分量都是 ε 的解析函数，并且 $\boldsymbol{x}(0) = \boldsymbol{x}$，$\boldsymbol{y}(0) = \boldsymbol{y}$.

在(6.7)式两端同时对 ε 在 $\varepsilon = 0$ 处取微商，有

$$B\boldsymbol{x}(0) + A\boldsymbol{x}'(0) = \lambda'(0)\boldsymbol{x}(0) + \lambda(0)\boldsymbol{x}'(0),$$

即

$$(B-\lambda'(0)I)\boldsymbol{x}(0)+(A-\lambda(0)I)\boldsymbol{x}'(0)=0.$$

利用 $\boldsymbol{y}(0)^H(A-\lambda(0)I)=0$ 及 $\boldsymbol{y}(0)^H\boldsymbol{x}(0)\neq 0$，得到

$$\boldsymbol{y}(0)^H(B-\lambda'(0)I)\boldsymbol{x}(0)=0$$

即

$$\lambda'(0)=\frac{\boldsymbol{y}(0)^H B\boldsymbol{x}(0)}{\boldsymbol{y}(0)^H\boldsymbol{x}(0)}=\frac{\boldsymbol{y}^H B\boldsymbol{x}}{\boldsymbol{y}^H\boldsymbol{x}}.$$

代入(6.6)式，便导出 $\lambda(\varepsilon)$ 的表达式(6.5)．□

(6.5)式表明：当 $|\varepsilon|\leqslant\varepsilon_0$ 时，$A+\varepsilon B$ 有一个特征值 $\lambda(\varepsilon)$，适合

$$|\lambda(\varepsilon)-\lambda|\leqslant\frac{\|\varepsilon B\|_2}{|s|}+O(|\varepsilon|^2),\quad s=\frac{\boldsymbol{y}^H\boldsymbol{x}}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2}. \tag{6.8}$$

例 6.1 $A=\begin{pmatrix}1 & 10^4\\ 0 & 2\end{pmatrix}$, $B=\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}$, $\varepsilon=10^{-6}$.

A 对应于单特征值 $\lambda_1=1$，有右特征向量 $\boldsymbol{x}_1=(1,\ 0)^T$ 和左特征向量 $\boldsymbol{y}_1^H=(1,-10^4)$； A 对应于单特征值 $\lambda_2=2$，有右特征向量 $\boldsymbol{x}_2=(10^4,1)^T$ 和左特征向量 $\boldsymbol{y}_2^H=(0,1)$．因此

$$|s_1|\approx 10^{-4},\quad |s_2|\approx 10^{-4}.$$

记 $A+\varepsilon B$ 的特征值为 $\lambda_1(\varepsilon)$ 和 $\lambda_2(\varepsilon)$．则由(6.8)可得估计式

$$|\lambda_i(\varepsilon)-\lambda_i|\lesssim\frac{\|\varepsilon B\|_2}{|s_i|}\approx 10^{-2},\quad i=1,2.$$

而实际计算表明：

$$\lambda_1(\varepsilon)\approx 1.0101021,\quad \lambda_2(\varepsilon)\approx 1.9898979,$$
$$|\lambda_1(\varepsilon)-\lambda_1|\approx 0.01,\quad |\lambda_2(\varepsilon)-\lambda_2|\approx 0.01.$$

6.2 几种条件数之间的关系

首先引证著名的 Канторович 不等式，作为预备．

定理 6.3 (Канторович 不等式)． 设 $A\in\mathbf{C}^{n\times n}$ 为正定阵，$\lambda(A)=\{\lambda_i\}$ 适合

$$M\geqslant\lambda_i\geqslant m>0,\quad i=1,\cdots,n.$$

则

$$x^H A x x^H A^{-1} x \leqslant \frac{(M+m)^2}{4Mm}(x^H x)^2, \ \forall x \in \mathbb{C}^n. \tag{6.9}$$

证明:

显然只需证明 $x \neq 0$ 和 $M > m$ 的情形.

利用 A 的酉-对角分解: $A = U^H \Lambda U$, U 为酉阵, $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, 代入(6.9)式左端,化为证明不等式

$$(\lambda_1 \xi_1 + \cdots + \lambda_n \xi_n)\left(\frac{\xi_1}{\lambda_1} + \cdots + \frac{\xi_n}{\lambda_n}\right) \leqslant \frac{(M+m)^2}{4Mm}, \tag{6.10}$$

其中 $\xi_1 + \cdots + \xi_n = 1$, 并且 $\xi_i \geqslant 0$, $i = 1, \cdots, n$.

令

$$\lambda_i = Mp_i + mq_i, \ \frac{1}{\lambda_i} = \frac{p_i}{M} + \frac{q_i}{m}, \ i = 1, \cdots, n.$$

由此可知 $p_i, q_i \geqslant 0$, $i = 1, \cdots, n$; 并且由

$$1 = \lambda_i \cdot \frac{1}{\lambda_i} = (Mp_i + mq_i)\left(\frac{p_i}{M} + \frac{q_i}{m}\right)$$

$$= (p_i + q_i)^2 + \frac{p_i q_i (M-m)^2}{Mm}$$

可知 $p_i + q_i \leqslant 1$, $i = 1, \cdots, n$.

再令

$$p = \sum_{i=1}^{n} \xi_i p_i, \quad q = \sum_{i=1}^{n} \xi_i q_i,$$

可得

$$p + q = \sum_{i=1}^{n} \xi_i (p_i + q_i) \leqslant \sum_{i=1}^{n} \xi_i = 1.$$

于是

$$\sum_{i=1}^{n} \lambda_i \xi_i \sum_{i=1}^{n} \frac{\xi_i}{\lambda_i} = \sum_{i=1}^{n} (Mp_i + mq_i)\xi_i \sum_{i=1}^{n}\left(\frac{p_i}{M} + \frac{q_i}{m}\right)\xi_i$$

$$= (pM + qm)\left(\frac{p}{M} + \frac{q}{m}\right)$$

$$= (p+q)^2 + pq\cdot\frac{(M-m)^2}{Mm}$$

$$= (p+q)^2\left[1+\frac{4pq}{(p+q)^2}\cdot\frac{(M-m)^2}{4Mm}\right]$$

$$\leqslant 1+\frac{(M-m)^2}{4Mm} = \frac{(M+m)^2}{4Mm}. \ \square$$

以下讨论(6.2)所定义的 $\nu(A)$ 与 $\nu_F(A)$ (在本段的以下部分,简单地记之为 ν 与 ν_F) 以及 (6.4) 所定义的 $s_1,\cdots,s_n$ 之间的关系,主要参考 Smith[150].

定理 6.4. 若 $A\in\mathbf{C}^{n\times n}$ 的每个特征值都是单特征值，则有

$$\frac{1}{s_i}\leqslant\frac{1}{2}\left(\nu+\frac{1}{\nu}\right),\ i=1,\cdots,n \tag{6.11}$$

和

$$\nu_F=\sum_{i=1}^{n}\frac{1}{s_i}. \tag{6.12}$$

证明:

据题设，存在非奇异阵 Q，使得 $Q^{-1}AQ=\Lambda$, $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$, $\lambda_i\neq\lambda_j$ $(i\neq j)$. 这时，A 属于每个 λ_i 的右特征向量与左特征向量在容许相差一个数量因子的条件下是唯一的. 因此,对于 λ_i, 可以分别取 $\boldsymbol{x}_i=Q\boldsymbol{e}_i$ 与 $\boldsymbol{y}_i^H=\boldsymbol{e}_i^TQ^{-1}$ 作为相应的右特征向量与左特征向量,其中 $\boldsymbol{e}_i$ 是 $I^{(n)}$ 的第 i 列向量,$i=1,\cdots,n$. 于是有

$$s_i=\frac{\boldsymbol{y}_i^H\boldsymbol{x}_i}{\|\boldsymbol{x}_i\|_2\|\boldsymbol{y}_i\|_2}=\frac{1}{\|Q\boldsymbol{e}_i\|_2\|\boldsymbol{e}_i^TQ^{-1}\|_2}.$$

利用 Конторович 不等式，可得

$$\frac{1}{s_i}=(\boldsymbol{e}_i^TQ^HQ\boldsymbol{e}_i\boldsymbol{e}_i^T(Q^HQ)^{-1}\boldsymbol{e}_i)^{\frac{1}{2}}$$

$$\leqslant\left[\frac{(\|Q\|_2^2+\|Q^{-1}\|_2^{-2})^2}{4\|Q\|_2^2\|Q^{-1}\|_2^{-2}}(\boldsymbol{e}_i^T\boldsymbol{e}_i)\right]^{\frac{1}{2}}$$

$$=\frac{1}{2}\frac{\|Q\|_2^2+\|Q^{-1}\|_2^{-2}}{\|Q\|_2\|Q^{-1}\|_2^{-1}}$$

$$= \frac{1}{2}\left(\|Q\|_2\|Q^{-1}\|_2 + \frac{1}{\|Q\|_2\|Q^{-1}\|_2} \right).$$

由此立即导出不等式(6.11).

以下证明(6.12). 根据题设,

$$AX = X\Lambda,\ X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n),\ \|\boldsymbol{x}_i\|_2 = 1,\ i = 1, \cdots, n,$$
$$\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n),\ \lambda_i \neq \lambda_j\ (i \neq j).$$

令

$$Z^H = X^{-1} = \begin{pmatrix} \boldsymbol{z}_1^H \\ \vdots \\ \boldsymbol{z}_n^H \end{pmatrix},\ Y^H = \mathrm{diag}(\|\boldsymbol{z}_1\|_2, \cdots, \|\boldsymbol{z}_n\|_2)^{-1} Z^H,$$

则有

$$Y^H A = \Lambda Y^H,\ Y = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n),\ \|\boldsymbol{y}_i\|_2 = 1,\ i = 1, \cdots, n$$

和

$$Y^H X \Lambda = \Lambda Y^H X.$$

因而

$$Y^H X = S = \mathrm{diag}(s_1, \cdots, s_n) > 0,$$

所以

$$X^{-1} = S^{-1} Y^H.$$

据题设, A 属于每个特征值的右特征向量和左特征向量在容许相差一个数量因子的条件下唯一确定,因此,凡使得 $Q^{-1}AQ = \Lambda$ 的 Q, 必形如

$$Q = XR,\ R = \mathrm{diag}(r_1, \cdots, r_n).$$

由此可得

$$\|Q\|_F^2 = \sum_{i=1}^{n} |r_i|^2.$$

进而,由

$$Q^{-1} = R^{-1}X^{-1} = R^{-1}S^{-1}Y^H$$

可得

$$\|Q^{-1}\|_F^2 = \sum_{i=1}^{n} \frac{1}{|r_i|^2 s_i^2}.$$

因此

$$\|Q\|_F^2\|Q^{-1}\|_F^2 = \sum_{i=1}^n |r_i|^2 \sum_{i=1}^n \frac{1}{|r_i|^2 s_i^2} \geqslant \left(\sum_{i=1}^n \frac{1}{s_i}\right)^2. \quad \text{Cauchy}$$

另一方面,取 $r_i = \frac{1}{\sqrt{s_i}}$,可使上式为等式. 所以(6.12)式成立.□

定理 6.5. $n-2+\nu+\nu^{-1} \leqslant \nu_F \leqslant \frac{1}{2}n(\nu+\nu^{-1})$.

证明:

任取 (6.1) 式中的一个 Q, Q^HQ 的特征值记为 $\omega_1 \geqslant \cdots \geqslant \omega_n > 0$, 并令

$$\Omega = \mathrm{diag}(\omega_1, \cdots, \omega_n), \boldsymbol{e} = (1, \cdots, 1)^T.$$

则

$$\begin{aligned}\|Q\|_F^2\|Q^{-1}\|_F^2 &= (\omega_1 + \cdots + \omega_n)(\omega_1^{-1} + \cdots + \omega_n^{-1}) \\ &= \boldsymbol{e}^T\Omega\boldsymbol{e} \cdot \boldsymbol{e}^T\Omega^{-1}\boldsymbol{e}.\end{aligned} \tag{6.13}$$

利用 Канторович 不等式,得到

$$\begin{aligned}\boldsymbol{e}^T\Omega\boldsymbol{e} \cdot \boldsymbol{e}^T\Omega^{-1}\boldsymbol{e} &\leqslant \frac{(\omega_1+\omega_n)^2}{4\,\omega_1\omega_n}(\boldsymbol{e}^T\boldsymbol{e})^2 \\ &= \left(\frac{n}{2}\cdot\frac{\omega_1+\omega_n}{\sqrt{\omega_1\omega_n}}\right)^2 = \left[\frac{n}{2}\left(\sqrt{\frac{\omega_1}{\omega_n}} + \sqrt{\frac{\omega_n}{\omega_1}}\right)\right]^2.\end{aligned}$$

代入(6.13)式, 即得出 $\nu_F \leqslant \frac{n}{2}(\nu+\nu^{-1})$.

令

$$\sigma = (\omega_1\omega_n)^{-\frac{1}{2}}(\omega_2 + \cdots + \omega_{n-1})$$

和

$$\tau = (\omega_1\omega_n)^{\frac{1}{2}}(\omega_2^{-1} + \cdots + \omega_{n-1}^{-1}).$$

于是有

$$\begin{aligned}\|Q\|_F^2\|Q^{-1}\|_F^2 &= (\omega_1\omega_n)^{-\frac{1}{2}}(\omega_1 + \cdots + \omega_n)(\omega_1\omega_n)^{\frac{1}{2}} \\ &\quad \cdot (\omega_1^{-1} + \cdots + \omega_n^{-1}) \\ &= \left(\sqrt{\frac{\omega_1}{\omega_n}} + \sigma + \sqrt{\frac{\omega_n}{\omega_1}}\right)\left(\sqrt{\frac{\omega_n}{\omega_1}} + \tau + \sqrt{\frac{\omega_1}{\omega_n}}\right)\end{aligned}$$

$$\geqslant \left[\sqrt{\frac{\omega_1}{\omega_n}} + \sqrt{\frac{\omega_n}{\omega_1}} + (\sigma\tau)^{\frac{1}{2}}\right]^2. \tag{6.14}$$

再注意到 $\sigma\tau \geqslant (n-2)^2$，并由 $\sqrt{\frac{\omega_1}{\omega_n}} \geqslant \nu \geqslant 1$ 可知

$$\sqrt{\frac{\omega_1}{\omega_n}} + \sqrt{\frac{\omega_n}{\omega_1}} \geqslant \nu + \nu^{-1},$$

代入(6.14),即得 $\nu_F \geqslant n-2+\nu+\nu^{-1}$. □

定理 6.6. 如果 A 是可正规化阵,则

$$\nu^4 \geqslant 1 + \frac{1}{2}(\|A^HA - AA^H\|_F / \|A^2\|_F)^2. \tag{6.15}$$

证明:

分解 $A = Q\Lambda Q^{-1}$，$\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 于是有

$$\begin{aligned}\|A^HA\|_F &= \|Q^{-H}\bar{\Lambda}Q^HQ\Lambda Q^{-1}\|_F \leqslant \|Q^{-1}\|_2^2\|\bar{\Lambda}Q^HQ\Lambda\|_F \\ &\leqslant \|Q^{-1}\|_2^2\|Q\|_2^2\|\bar{\Lambda}\Lambda\|_F.\end{aligned}$$

将 $\|\bar{\Lambda}\Lambda\|_F = \|\Lambda^2\|_F = \|A^2\|_F$ 代入上式右端,得到

$$\|Q\|_2^2\|Q^{-1}\|_2^2 \geqslant \|A^2\|_F^{-1}\|A^HA\|_F,$$

从而

$$\nu^2 \geqslant \|A^2\|_F^{-1}\|A^HA\|_F.$$

再利用

$$\|A^HA - AA^H\|_F^2 = 2\|A^HA\|_F^2 - 2\|A^2\|_F^2,$$

便得出

$$\begin{aligned}\nu^2 &\geqslant \|A^2\|_F^{-1}\left(\|A^2\|_F^2 + \frac{1}{2}\|A^HA - AA^H\|_F^2\right)^{\frac{1}{2}} \\ &= \left[1 + \frac{1}{2}\left(\frac{\|A^HA - AA^H\|_F}{\|A^2\|_F}\right)^2\right]^{\frac{1}{2}}. \quad \square\end{aligned}$$

注 6.1 对非奇异阵 Q 乘以适当的非奇异阵(一般考虑对角阵或块对角阵),使得 $\|Q\|_2\|Q^{-1}\|_2$ 变小的手续,通常叫做平衡,或叫做预处理. 例如对于例 6.1 所示的矩阵 $A = \begin{pmatrix} 1 & 10^4 \\ 0 & 2 \end{pmatrix}$,存在 $Q=$

$\begin{pmatrix} 1 & 10^4 \\ 0 & 1 \end{pmatrix}$，使得

$$Q^{-1}AQ = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

计算可知，$\|Q\|_2\|Q^{-1}\|_2 \approx 10^8$. 如果对 Q 右乘以对角阵 $D = \mathrm{diag}(1, 10^{-4})$，同样有

$$(QD)^{-1}A(QD) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},$$

但这时

$$\|QD\|_2\|(QD)^{-1}\|_2 = \left\|\begin{pmatrix} 1 & 1 \\ 0 & 10^{-4} \end{pmatrix}\right\|_2 \left\|\begin{pmatrix} 1 & -10^4 \\ 0 & 10^4 \end{pmatrix}\right\|_2$$
$$\approx 10^4 \ll 10^8.$$

如何通过相似变换将 $\|Q\|_2\|Q^{-1}\|_2$ 尽可能地缩小，以及如何估算 $\|Q\|_2\|Q^{-1}\|_2$ 的比较靠近准确值的上界，是一个有待进一步研究的问题. 读者可参看文献[20]、[21]、[33]、[45]、[46]、[49]、[67]、[82]、[91]、[123]、[129]、[149]、[150]和[170].

习题

1. 设 $A = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n) \in \mathbf{C}^{n\times n}$ 为非奇异阵. 试证：如果

$$|\boldsymbol{a}_1^H\boldsymbol{a}_2| \geqslant \|\boldsymbol{a}_1\|_2\|\boldsymbol{a}_2\|_2(1-\varepsilon), \quad 0 < \varepsilon < 1$$

(此示二向量 $\boldsymbol{a}_1$ 与 $\boldsymbol{a}_2$ 之间的夹角 θ 满足 $1-\varepsilon \leqslant |\cos\theta| \leqslant 1$)，则

$$\|A\|_2\|A^{-1}\|_2 \geqslant \frac{1}{\sqrt{\varepsilon}}.$$

2. 设 $A = \begin{pmatrix} I & B \\ B^H & I \end{pmatrix}$，$\|B\|_2 < 1$. 试证

$$\|A\|_2\|A^{-1}\|_2 = \frac{1+\|B\|_2}{1-\|B\|_2}.$$

3. 设 $A, B \in \mathbf{C}^{n\times n}$，其中 B 为可正规化阵，$\lambda(B) = \{\mu_i\}$. 令

$$\mathscr{D}_i = \{\lambda \in \mathbf{C}: |\lambda - \mu_i| \leqslant \nu(B)\|A - B\|_2\},$$
$$i = 1, \cdots, n,$$

其中 $\nu(B)$ 如(6.2)定义．则

$$\lambda(A) \subseteq \bigcup_{i=1}^{n} \mathscr{D}_i.$$

4．令 $L_i(\boldsymbol{l}_i) \equiv E(\boldsymbol{l}_i, \boldsymbol{e}_i; 1)$ 表示初等下三角阵（见第一章§2），其中

$$\boldsymbol{l}_i = (0, \cdots, 0, l_{i+1,i}, \cdots, l_{ni})^T,$$

$\boldsymbol{e}_i$ 是 $I^{(n)}$ 的第 i 列向量．试证

$$\|L_i(\boldsymbol{l}_i)\|_2 \|L_i(\boldsymbol{l}_i)^{-1}\|_2 = 1 + \frac{\rho(\rho + \sqrt{\rho^2 + 4})}{2},$$

其中 $\rho = \|\boldsymbol{l}_i\|_2$.

§7　特征空间的扰动界限

本节讨论 Hermite 阵．

7.1　Rayleigh 商和剩余

定义 7.1. 设 $A \in \mathbb{C}^{n \times n}$ 为 Hermite 阵．A的 Rayleigh 商，是指在 $\mathbb{C}^n$ 上定义的一个实值函数

$$\rho(\boldsymbol{x}) \equiv \rho(\boldsymbol{x}; A) = \frac{\boldsymbol{x}^H A \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{x}}, \quad \boldsymbol{x} \in \mathbb{C}^n, \ \boldsymbol{x} \neq 0. \tag{7.1}$$

A给定后，$\rho(\boldsymbol{x})$ 又叫做 $\boldsymbol{x}$ 的 Rayleigh 商．

Rayleigh 商的性质：

1）齐次：$\rho(\alpha\boldsymbol{x}) = \rho(\boldsymbol{x})$, $\alpha \in \mathbb{C}$, $\alpha \neq 0$.

2）有界：若A的特征值为 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$，则

$$\lambda_n \leqslant \rho(\boldsymbol{x}) \leqslant \lambda_1, \quad \forall \boldsymbol{x} \in \mathbb{C}^n, \ \boldsymbol{x} \neq 0.$$

3）极小剩余：对于每个 $\boldsymbol{x} \in \mathbb{C}^n$，有

$$\|A\boldsymbol{x} - \rho(\boldsymbol{x})\boldsymbol{x}\|_2 = \inf_{\mu \in \mathbb{C}} \|A\boldsymbol{x} - \mu\boldsymbol{x}\|_2.$$

事实上，由

$$A\boldsymbol{x} - \mu\boldsymbol{x} = A\boldsymbol{x} - \rho(\boldsymbol{x})\boldsymbol{x} + (\rho(\boldsymbol{x}) - \mu)\boldsymbol{x}$$

及

$$x^H(Ax-\rho(x)x)=0$$

可知

$$\begin{aligned}\|Ax-\mu x\|_2^2 &= \|Ax-\rho(x)x\|_2^2+\|(\rho(x)-\mu)x\|_2^2\\ &\geqslant \|Ax-\rho(x)x\|_2^2,\end{aligned}$$

并且当且仅当 $\mu=\rho(x)$ 时,可以达到等式.

4) 作为特征值的近似: 定理 4.1 表明, 如果 A 的特征值为 $\lambda_1\geqslant\cdots\geqslant\lambda_n$, 并任取 $x\in C^n$, $\|x\|_2=1$, 则有

$$\min_{1\leqslant i\leqslant n}|\lambda_i-\rho(x)|\leqslant\|Ax-\rho(x)x\|_2. \tag{7.2}$$

从定义 7.1 和估计式(7.2)可以看出: 如果 x 是 Hermite 阵 A 的特征向量,则 $\rho(x)$ 是 A 的特征值; 如果 x 是 A 的一个近似特征向量,则 $\rho(x)$ 是 A 的一个近似特征值. 下述定理表明, $\rho(x)$ 作为 A 的近似特征值的精确度, 高于 x 作为 A 的近似特征向量的精确度.

定理 7.1. 设 x_1 是 n 阶 Hermite 阵 A 属于特征值 λ_1 的特征向量. 如果 $x\in C^n$, 适合

$$\sin\theta(R(x),R(x_1))=O(\varepsilon), \tag{7.3}$$

则

$$\rho(x)-\lambda_1=O(\varepsilon^2).$$

(7.3)式中的 $\theta(R(x),R(x_1))$, 见第二章(4.1).

证明:

首先指出, 如果 $\sin\theta(R(x),R(x_1))=0$, 则定理结论显然成立. 因此只需研究 x 与 x_1 不共线的情形.

其次, 无妨假设 $\|x_1\|_2=\|x\|_2=1$, 因为这并不影响 $\sin\theta(R(x),R(x_1))$ 与 $\rho(x)$ 之值.

根据 $\sin\theta(R(x),R(x_1))$ 的定义,如果令

$$x=x_1+\delta u,\ u\in C^n,\ \|u\|_2=1,\ \delta>0,$$

则由(7.3)得到

$$\delta\sqrt{1-x_1^Huu^Hx_1}=O(\varepsilon).$$

再由 $|\boldsymbol{x}_1^H\boldsymbol{u}|<1$ 得知 $\delta=O(\varepsilon)$，因而

$$\boldsymbol{x}=\boldsymbol{x}_1+O(\varepsilon). \tag{7.3$'$}$$

设 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 是由 A 的特征向量所构成的标准正交基. 令 $X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)$ 和 $\boldsymbol{x}=X\boldsymbol{y}$，其中 $\boldsymbol{y}=(\eta_1,\cdots,\eta_n)^T$. 于是

$$\rho(\boldsymbol{x})\equiv\boldsymbol{x}^HA\boldsymbol{x}=\boldsymbol{y}^HX^HAX\boldsymbol{y}=|\eta_1|^2\lambda_1+\sum_{i=2}^n|\eta_i|^2\lambda_i.$$

其中 $\lambda_1,\cdots,\lambda_n$ 是 A 的特征值.

由 (7.3)$'$和 $\boldsymbol{x}=\eta_1\boldsymbol{x}_1+\cdots+\eta_n\boldsymbol{x}_n$ 可推知

$$\eta_1=1+O(\varepsilon),\ \eta_i=O(\varepsilon),\ i=2,\cdots,n.$$

因此

$$\begin{aligned}\rho(\boldsymbol{x})&=|\eta_1|^2\lambda_1+\sum_{i=2}^n|\eta_i|^2\lambda_i\\&=\left(1-\sum_{i=2}^n|\eta_i|^2\right)\lambda_1+\sum_{i=2}^n|\eta_i|^2\lambda_i\\&=\lambda_1+O(\varepsilon^2).\ \square\end{aligned}$$

注 7.1. 任意二向量 $\boldsymbol{x}$ 与 $\boldsymbol{x}_1\in\mathbb{C}^n$ (不限定 $\|\boldsymbol{x}\|_2=\|\boldsymbol{x}_1\|_2=1$)，如果适合关系式 (7.3)$'$，则必适合条件 (7.3) (利用 $\sin\theta(R(\boldsymbol{x}),R(\boldsymbol{x}_1))$ 的定义即可证明)，因而可以利用定理 7.1 的结论. 例如

$$A=\begin{pmatrix}1.005 & -0.9950\\ -0.9950 & 1.005\end{pmatrix}.$$

A 对应于特征值 $\lambda_1=2$ 的特征向量为 $\boldsymbol{x}_1=(1,-1)^T$. 今取 $\boldsymbol{x}=(1.01,-0.99)^T$，有 $\rho(\boldsymbol{x})=1.9998$，精确到四位数字. 在这个例子里，$\boldsymbol{x}=\boldsymbol{x}_1+(0.1,0.01)^T$，$\rho(\boldsymbol{x})-\lambda_1=-0.0002$.

通过 A 的 Rayleigh 商 $\rho(\boldsymbol{x})$ 构造剩余 $A\boldsymbol{x}-\rho(\boldsymbol{x})\boldsymbol{x}$，不仅可以用来界定 $\rho(\boldsymbol{x})$ 与 A 的特征值集合的距离(见(7.2))，而且可以用它界定 $\boldsymbol{x}$ 与 A 的特征向量的距离. 具体说来，下述结论成立(参考[136]).

定理 7.2. 设 $A\in\mathbb{C}^{n\times n}$ 为 Hermite 阵，$\boldsymbol{y}\in\mathbb{C}^n$ 是一单位向量. 令 $\mu=\rho(\boldsymbol{y})$ 和

$$r(\boldsymbol{y}) \equiv A\boldsymbol{y} - \rho(\boldsymbol{y})\boldsymbol{y} = (A - \mu I)\boldsymbol{y}. \tag{7.4}$$

如果A的特征值中距μ最近者是λ，A属于λ的单位特征向量为$\boldsymbol{x}$，并且

$$\delta \equiv \min_{\substack{\lambda_i \neq \lambda \\ \lambda_i \in \lambda(A)}} |\lambda_i - \mu| > 0,$$

则有

$$\sin\theta(R(\boldsymbol{x}), R(\boldsymbol{y})) \leqslant \frac{\|r(\boldsymbol{y})\|_2}{\delta} \tag{7.5}$$

和

$$|\lambda - \mu| \leqslant \frac{\|r(\boldsymbol{y})\|_2^2}{\delta}. \tag{7.6}$$

证明:

以下用θ代表 $\theta(R(\boldsymbol{x}), R(\boldsymbol{y})) \equiv \arccos|\boldsymbol{x}^H\boldsymbol{y}|$.

因可用一适当的模为1的复数乘以$\boldsymbol{x}$，故无妨假设 $\boldsymbol{x}^H\boldsymbol{y} = \cos\theta$. 然后取一单位向量$\boldsymbol{w}$，满足 $\boldsymbol{x}^H\boldsymbol{w} = 0$ 和 $\boldsymbol{y}^H\boldsymbol{w} = \sin\theta$. 于是有

$$\boldsymbol{y} = \boldsymbol{x}\cos\theta + \boldsymbol{w}\sin\theta. \tag{7.7}$$

从而

$$\begin{aligned} r(\boldsymbol{y}) &= (A - \mu I)\boldsymbol{x}\cos\theta + (A - \mu I)\boldsymbol{w}\sin\theta \\ &= (\lambda - \mu)\boldsymbol{x}\cos\theta + (A - \mu I)\boldsymbol{w}\sin\theta. \end{aligned} \tag{7.8}$$

注意到 $\boldsymbol{x}^H(A - \mu I)\boldsymbol{w} = 0$，因此有

$$\|r(\boldsymbol{y})\|_2^2 = (\lambda - \mu)^2\cos^2\theta + \|(A - \mu I)\boldsymbol{w}\|_2^2\sin^2\theta. \tag{7.9}$$

取A的特征向量所组成的标准正交基 $\boldsymbol{x}, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，构造酉阵 $X = (\boldsymbol{x}, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$. 有

$$A = X\Lambda X^H,\ \Lambda = \operatorname{diag}(\lambda, \lambda_2, \cdots, \lambda_n),$$

$$X^H\boldsymbol{w} = (0, \omega_2, \cdots, \omega_n)^T,\ \sum_{i=2}^{n} |\omega_i|^2 = 1$$

和

$$\begin{aligned} \|(A - \mu I)\boldsymbol{w}\|_2^2 &= \|X(\Lambda - \mu I)X^H\boldsymbol{w}\|_2^2 \\ &= \sum_{i=2}^{n} |\lambda_i - \mu|^2|\omega_i|^2 \geqslant \delta^2. \end{aligned}$$

代入(7.9),即得(7.5).

再由(7.4)、(7.7)和(7.8)可知

$$0 = \boldsymbol{y}^H r(\boldsymbol{y}) = (\lambda - \mu)\cos^2\theta + \boldsymbol{w}^H(A - \mu I)\boldsymbol{w}\sin^2\theta,$$

由此可得出

$$\cos^2\theta = -\frac{\boldsymbol{w}^H(A - \mu I)\boldsymbol{w}}{\boldsymbol{w}^H(\lambda I - A)\boldsymbol{w}}, \quad \sin^2\theta = \frac{\lambda - \mu}{\boldsymbol{w}^H(\lambda I - A)\boldsymbol{w}},$$

代入(7.9),

$$\|r(\boldsymbol{y})\|_2^2 = \frac{(\mu - \lambda)^2\boldsymbol{w}^H(A - \mu I)\boldsymbol{w} + \|(A - \mu I)\boldsymbol{w}\|_2^2(\mu - \lambda)}{\boldsymbol{w}^H(A - \lambda I)\boldsymbol{w}}. \tag{7.10}$$

计算可知,上式右端的分子为 $(\mu - \lambda)\boldsymbol{w}^H(A - \mu I)(A - \lambda I)\boldsymbol{w}$ 而

$$\boldsymbol{w}^H(A - \mu I)(A - \lambda I)\boldsymbol{w} = \sum_{i=2}^{n}(\lambda_i - \mu)(\lambda_i - \lambda)|\omega_i|^2,$$

其中

$$(\lambda_i - \mu)(\lambda_i - \lambda) = |\lambda_i - \mu||\lambda_i - \lambda| \geqslant \delta|\lambda_i - \lambda|, \quad i = 2, \cdots, n$$

(根据 $|\lambda - \mu| \leqslant \delta \leqslant |\lambda_i - \mu|, i = 2, \cdots, n$). 代入(7.10),

$$\|r(\boldsymbol{y})\|_2^2 \geqslant \frac{(\mu - \lambda)\delta\sum_{i=2}^{n}|\lambda_i - \lambda||\omega_i|^2}{\sum_{i=2}^{n}(\lambda_i - \lambda)|\omega_i|^2}$$

$$= \frac{|\mu - \lambda|\delta\sum_{i=2}^{n}|\lambda_i - \lambda||\omega_i|^2}{\left|\sum_{i=2}^{n}(\lambda_i - \lambda)|\omega_i|^2\right|},$$

因此

$$|\mu-\lambda| \leqslant \frac{\|r(y)\|_2^2}{\delta} \cdot \left|\frac{\sum_{i=2}^{n}(\lambda_i-\lambda)|\omega_i|^2}{\sum_{i=2}^{n}|\lambda_i-\lambda||\omega_i|^2}\right| \leqslant \frac{\|r(y)\|_2^2}{\delta}. \quad \square$$

上面所讨论的 Rayleigh 商，是指在 $\mathbf{C}^n$ 上所定义的一个实值函数． 对于 Hermite 阵 $A\in\mathbf{C}^{n\times n}$，也可以在 $\mathbf{C}^{n\times l}$ $(1<l\leqslant n)$ 上定义 Rayleigh 商．

定义 7.2. 设 $A\in\mathbf{C}^{n\times n}$ 是 Hermite 阵． A 在 $\mathbf{C}^{n\times l}$ $(1\leqslant l\leqslant n)$ 上的 Rayleigh 商是指

$$\rho(Q_1)\equiv Q_1^H A Q_1,$$

其中 $Q_1\in\mathbf{C}^{n\times l}$，$Q_1^H Q_1=I$．

A 给定后，$\rho(Q_1)$ 又叫做 Q_1 的 Rayleigh 商．

利用定理 3.2，可以立即得知 Rayleigh 商 $\rho(Q_1)$ 的有界性，即：若 A 的特征值为 $\lambda_1\geqslant\cdots\geqslant\lambda_n$，则

$$\lambda_n I\leqslant\rho(Q_1)\leqslant\lambda_1 I. \quad \forall Q_1\in\mathbf{C}^{n\times l},\ Q_1^H Q_1=I.$$

Rayleigh 商的性质 3)，可以推广为下述定理．

定理 7.3. 设 $A\in\mathbf{C}^{n\times n}$ 为 Hermite 阵，$Q_1\in\mathbf{C}^{n\times l}$ $(1\leqslant l\leqslant n)$，$Q_1^H Q_1=I$． 令 $H_1=Q_1^H A Q_1$． 则

$$\|AQ_1-Q_1H_1\|_2=\inf_{M\in\mathbf{C}^{l\times l}}\|AQ_1-Q_1M\|_2.$$

证明:

记 $R=AQ_1-Q_1H_1$，$R_M=AQ_1-Q_1M$． 有

$$R_M=R+S,\quad S=Q_1(H_1-M).$$

容易验证 $S^HR=0$． 因此

$$R_M^H R_M=R^HR+S^HS\geqslant R^HR.$$

从而

$$\|R\|_2\leqslant\|R_M\|_2,\ \forall M\in\mathbf{C}^{l\times l}.$$

另一方面，取 $M=H_1$，可使上式中的等式成立． $\square$

作为 Rayleigh 商性质 4) 的推广，Kahan 证明了下列两条定理(定理 7.4 和定理 7.5．参考[136])．

定理 7.4. 设 $A\in\mathbf{C}^{n\times n}$ 为 Hermite 阵，$Q_1\in\mathbf{C}^{n\times l}$ $(1\leqslant l\leqslant n)$，$Q_1^HQ_1=I$. 令 $H_1=Q_1^HAQ_1$ 和 $R=AQ_1-Q_1H_1$. $\lambda(H_1)=\{\mu_i\}_{i=1}^l$. 则存在 $\lambda_{i'}\in\lambda(A)$，$i'=1,\cdots,l$，使得

$$|\lambda_{i'}-\mu_i|\leqslant\|R\|_2,\quad i=1,\cdots,l. \tag{7.11}$$

证明:

首先取 $Q_2\in\mathbf{C}^{n\times(n-l)}$，使 $Q=(Q_1,Q_2)$ 为一酉阵. 于是

$$\begin{aligned}Q^HAQ&=\begin{pmatrix}Q_1^HAQ_1 & Q_1^HAQ_2\\ Q_2^HAQ_1 & Q_2^HAQ_2\end{pmatrix}=\begin{pmatrix}H_1 & B^H\\ B & H_2\end{pmatrix}\\&=\begin{pmatrix}H_1 & 0\\ 0 & H_2\end{pmatrix}+\begin{pmatrix}0 & B^H\\ B & 0\end{pmatrix}.\end{aligned} \tag{7.12}$$

另一方面，有

$$\begin{aligned}Q^HR&=Q^HAQ_1-Q^HQ_1H_1\\&=Q^HAQ\cdot Q^HQ_1-Q^HQ_1\cdot H_1\\&=\begin{pmatrix}H_1 & B^H\\ B & H_2\end{pmatrix}\begin{pmatrix}I\\0\end{pmatrix}-\begin{pmatrix}I\\0\end{pmatrix}H_1=\begin{pmatrix}0\\B\end{pmatrix}.\end{aligned} \tag{7.13}$$

根据定理 3.6，由(7.12)可知，存在 $\lambda_i\in\lambda(A)$，使得

$$|\lambda_i-\mu_i|\leqslant\left\|\begin{pmatrix}0 & B^H\\ B & 0\end{pmatrix}\right\|_2=\|B\|_2,\quad i=1,\cdots,l. \tag{7.14}$$

再由(7.13)知

$$\|B\|_2=\left\|\begin{pmatrix}0\\B\end{pmatrix}\right\|_2=\|Q^HR\|_2=\|R\|_2.$$

代入(7.14)即得到不等式(7.11). □

定理 7.5. 在定理 7.4 的假设条件下，存在 $\lambda_{i''}\in\lambda(A)$，$i=1,\cdots,l$，使得

$$\sqrt{\sum_{i=1}^{l}|\lambda_{i''}-\mu_i|^2}\leqslant\sqrt{2}\,\|R\|_F. \tag{7.15}$$

证明:

利用 Hoffman-Wielandt 定理，从(7.12)和(7.13)可立即导出不等式(7.15). □

作为定理 7.2 的推广，有下述结论.

定理 7.6. 设 $A\in\mathbb{C}^{n\times n}$ 为 Hermite 阵，$Q_1\in\mathbb{C}^{n\times l}$ $(1\leqslant l\leqslant n)$，$Q_1^HQ_1=I$. 令 $H_1=Q_1^HAQ_1$ 和 $R=AQ_1-Q_1H_1$. $\lambda(A)=\{\lambda_i\}$，$\lambda(H_1)=\{\mu_i\}_{i=1}^l$. 如果 A 的特征值 $\lambda_{1'},\cdots,\lambda_{l'}$ 满足不等式(7.11)，A 属于 $\lambda_{1'},\cdots,\lambda_{l'}$ 的特征向量所张成的子空间为 $R(Z_1)$，$Z_1\in\mathbb{C}_l^{n\times l}$；并且如果对于

$$\mathscr{I}=\{1,\cdots,n\}\backslash\{1',\cdots,l'\},$$

有

$$\delta\equiv\min_{\substack{i\in\mathscr{I}\\ j=1,\cdots,l}}|\lambda_i-\mu_j|>0, \tag{7.16}$$

则

$$\|\sin\Theta(R(Z_1),R(Q_1))\|_F\leqslant\frac{\|R\|_F}{\delta}, \tag{7.17}$$

其中 $\Theta(R(Z_1),R(Q_1))$ 的定义见第二章(4.3)—(4.4).

定理 7.6 可以从下一段的 Davis-Kahan $\sin\theta$ 第二定理导出，证明放在定理 7.10 之后.

7.2 Davis-Kahan $\sin\theta$ 定理

定义 7.3. 设 $A\in\mathbb{C}^{n\times n}$ 为 Hermite 阵. $\mathbb{C}^n$ 内的子空间 $\mathscr{X}$ 叫做 A 的特征空间，如果

$$A\mathscr{X}\subseteq\mathscr{X}. \tag{7.18}$$

A 的特征向量的任一集合，显然张成 A 的一个特征空间. 反之，A 的任一特征空间 $\mathscr{X}$ 必由其特征向量张成. 这是因为，首先可从 $\mathscr{X}$ 中选出一组标准正交基 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_l$，记 $X_1=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_l)$. 据(7.18)，应有 $AX_1=X_1A_1$，其中 $A_1\in\mathbb{C}^{l\times l}$ 为 Hermite 阵，可分解为 $A_1=U_1\Lambda_1U_1^H$，$\Lambda_1=\mathrm{diag}(\lambda_1,\cdots,\lambda_l)$，$U_1\in\mathbb{C}^{l\times l}$ 为酉阵. 于是

$$A(X_1U_1)=(X_1U_1)\Lambda_1.$$

这表明 X_1U_1 的列向量是 A 的特征向量，并且有

$$\mathscr{X}=R(X_1)=R(X_1U_1).$$

因此，Hermite 阵的特征空间可以等价地定义如下.

定义 7.4. 设 $A\in\mathbf{C}^{n\times n}$ 为 Hermite 阵. 如果 $\mathscr{X}=R(X_1)$, 其中 $X_1\in\mathbf{C}_l^{n\times l}$ $(1\leqslant l\leqslant n)$, 并且存在 $A_1\in\mathbf{C}^{l\times l}$, 使得

$$AX_1=X_1A_1, \tag{7.19}$$

则称 $\mathscr{X}$ 为 A 的一个 l 维特征空间.

Davis-Kahan $\sin\theta$ 定理是关于 Hermite 阵特征空间扰动的著名结果. 它假定两个 Hermite 阵 A 与 $A+E$ 对应的特征空间已经存在,并对 A 与 $A+E$ 的特征值的分布限定一些条件,然后在这些条件之下,求出两个对应特征空间距离的上界.

在引证 Davis-Kahan $\sin\theta$ 定理之前,还须再对矩阵的酉不变范数作些补充.

定义 7.5. 设 $\|\ \|$ 是 $\mathbf{C}^{l\times l}$ 上的酉不变范数, Φ 是 $\mathbf{R}^l$ 上与 $\|\ \|$相对应的 SG 函数. 又设自然数 m 与 k 满足 $\min\{m,k\}\leqslant l$. 所谓范数 $\|\ \|$ 在 $\mathbf{C}^{m\times k}$ 上的延拓 (仍记为 $\|\ \|$), 是指如下定义的非负实值函数: 对任一 $A\in\mathbf{C}^{m\times k}$, 如果它的奇异值分解是

$$A=U\begin{pmatrix}\sigma_1 & & \\ & \ddots & \sigma_k \\ & 0 & \end{pmatrix}V^H \quad \text{或} \quad A=U\begin{pmatrix}\sigma_1 & & \\ & \ddots & 0 \\ & \sigma_m & \end{pmatrix}V^H,$$

其中 U 与 V 为酉阵, $\sigma_1,\sigma_2,\cdots\geqslant 0$, 则定义

$$\|A\|=\Phi(\sigma_1,\cdots,\sigma_p,0,\cdots,0),\ p=\min\{m,k\}.$$

显然,酉不变范数的延拓仍是酉不变范数.

利用范数的延拓,可得下述引理.

引理 7.1. 设 $X_1,\ \tilde{X}_1\in\mathbf{C}^{n\times l}(1\leqslant l\leqslant n-1)$, 并且 $X_1^HX_1=\tilde{X}_1^H\tilde{X}_1=I$. 又设 $\tilde{X}=(\tilde{X}_1,\tilde{X}_2)\in\mathbf{C}^{n\times n}$ 是酉阵. 令 $\mathscr{X}=R(X_1)$ 和 $\tilde{\mathscr{X}}=R(\tilde{X}_1)$. 则对于 $\mathbf{C}^{l\times l}$ 上的任一酉不变范数 $\|\ \|$, 有

$$\|\sin\Theta(\mathscr{X},\tilde{\mathscr{X}})\|=\|\tilde{X}_2^HX_1\|. \tag{7.20}$$

上式右端的 $\|\ \|$, 是 $\mathbf{C}^{l\times l}$ 上的范数 $\|\ \|$ 在 $\mathbf{C}^{(n-l)\times l}$ 上的延拓.

证明:

因为 $\tilde{X}$ 是酉阵, 所以根据 $\Theta(\mathscr{X},\tilde{\mathscr{X}})$ 的定义 (见第二章 (4.3)—(4.4)),

$$\sin^2\Theta(\mathscr{X},\widetilde{\mathscr{X}}) = I - X_1^H\widetilde{X}_1\widetilde{X}_1^H X_1 = X_1^H\widetilde{X}_2\widetilde{X}_2^H X_1.$$

设 $\widetilde{X}_2^H X_1$ 的奇异值分解为 $\widetilde{X}_2^H X_1 = U\Sigma V^H$，其中 U 与 V 为酉阵，并且

$$\Sigma = \begin{cases} \begin{pmatrix} \Sigma_1 \\ 0 \end{pmatrix} \begin{matrix} l \\ n-2l \end{matrix}, \ \Sigma_1 = \mathrm{diag}(\sigma_1, \cdots, \sigma_l),\ 2l \leqslant n \\ \quad\ \ l \\ \underset{n-l\ \ 2l-n}{(\Sigma_1', 0)}\ n-l, \Sigma_1' = \mathrm{diag}(\sigma_1, \cdots, \sigma_{n-l}),\ 2l \geqslant n. \end{cases} \tag{7.21}$$

于是有

$$\sin\Theta(\mathscr{X}, \widetilde{\mathscr{X}}) = \begin{cases} V\Sigma_1 V^H, & 2l \leqslant n \\ V\begin{pmatrix} \Sigma_1' & 0 \\ 0 & 0 \end{pmatrix} V^H, & 2l \geqslant n. \end{cases} \tag{7.22}$$

据定义 7.5，从(7.21)和 (7.22) 立即得到(7.20)。□

引理 7.2. 设有矩阵方程

$$BX - XA = C,\ X \in \mathbf{C}^{m\times n}.$$

其中 $A \in \mathbf{C}^{n\times n}$，$B \in \mathbf{C}_m^{m\times m}$，$C \in \mathbf{C}^{m\times n}$. 如果存在 $\alpha \geqslant 0$ 和 $\delta > 0$，使得

$$\|A\|_2 \leqslant \alpha,\ \|B^{-1}\|_2 \leqslant \frac{1}{\alpha+\delta},$$

则对于 $\mathbf{C}^{m\times n}$ 上的任一种酉不变范数 $\|\ \ \|$，有

$$\|X\| \leqslant \frac{\|C\|}{\delta}.$$

证明:

根据第二章定理 3.4 和题设条件，有

$$\|X\| \leqslant \|B^{-1}\|_2 \|BX\| \leqslant \frac{\|BX\|}{\alpha+\delta},$$

因而

$$\|BX\| \geqslant (\alpha+\delta)\|X\|;$$

另一方面，有

$$\|XA\| \leqslant \|A\|_2\|X\| \leqslant \alpha\|X\|.$$

因此

$$\|C\| \geqslant \|BX\| - \|AX\| \geqslant \delta\|X\|. \quad \square$$

定理 7.7 ($\sin\theta$ 第一定理)[75]. 设 $A, \widetilde{A} = A + E \in \mathbf{C}^{n\times n}$ 皆为 Hermite 阵，并假定 $X = (X_1, X_2)$ 与 $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2)$ 为 n 阶酉阵，X_1 与 $\widetilde{X}_1 \in \mathbf{C}^{n\times l}$ $(1 \leqslant l \leqslant n-1)$，使得

$$X^H A X = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}, \ \widetilde{X}^H \widetilde{A} \widetilde{X} = \begin{pmatrix} \widetilde{A}_{11} & 0 \\ 0 & \widetilde{A}_{22} \end{pmatrix}, \tag{7.23}$$

其中 $A_{11}, \widetilde{A}_{11} \in \mathbf{C}^{l\times l}$. 令 $\mathscr{X} = R(X_1)$，$\widetilde{\mathscr{X}} = R(\widetilde{X}_1)$ 和

$$R = \widetilde{A} X_1 - X_1 A_{11} \tag{7.24}$$

以及

$$\Delta = [\alpha, \beta] \subset \mathbf{R}, \Delta' = \mathbf{R} \backslash (\alpha - \delta, \beta + \delta), \delta > 0. \tag{7.25}$$

如果

$$\lambda(A_{11}) \subset \Delta, \ \lambda(\widetilde{A}_{22}) \subset \Delta' \tag{7.26}$$

(或者 $\lambda(A_{11}) \subset \Delta'$，$\lambda(\widetilde{A}_{22}) \subset \Delta$)，则对于 $\mathbf{C}^{l\times l}$ 上的任一酉不变范数 $\|\ \ \|$，有

$$\|\sin\Theta(\mathscr{X}, \widetilde{\mathscr{X}})\| \leqslant \frac{\|R\|}{\delta}. \tag{7.27}$$

上式右端所使用的范数$\|\ \ \|$是 $\mathbf{C}^{l\times l}$ 上的范数$\|\ \ \|$在 $\mathbf{C}^{n\times l}$ 上的延拓.

证明:

因为同时用 $A - \dfrac{\alpha+\beta}{2} I$ 与 $\widetilde{A} - \dfrac{\alpha+\beta}{2} I$ 代替 A 与 $\widetilde{A}$，并不影响 R 和定理的条件，所以无妨假设 $0 \leqslant \alpha = -\beta$.

利用分解式 (7.23)，有

$$R = \widetilde{X}_1 \widetilde{A}_{11} \widetilde{X}_1^H X_1 + \widetilde{X}_2 \widetilde{A}_{22} \widetilde{X}_2^H X_1 - X_1 A_{11},$$

因而

$$\widetilde{X}_2^H R = \widetilde{A}_{22}(\widetilde{X}_2^H X_1) - (\widetilde{X}_2^H X_1) A_{11}. \tag{7.28}$$

令 $Z = \widetilde{X}_2^H X_1 \in \mathbf{C}^{(n-l)\times l}$，它满足方程

$$\widetilde{A}_{22} Z - Z A_{11} = \widetilde{X}_2^H R, \tag{7.29}$$

其中 $\|\tilde{A}_{22}^{-1}\|_2 \geqslant \frac{1}{\alpha+\delta}$，$\|A_{11}\|_2 \leqslant \alpha$ (据(7.26))．因此根据引理 7.2，以及 $(\tilde{X}_2^H R)^H(\tilde{X}_2^H R) \leqslant R^H R$，有

$$\|Z\| \leqslant \frac{\|\tilde{X}_2^H R\|}{\delta} \leqslant \frac{\|R\|}{\delta}. \tag{7.30}$$

再根据引理 7.1，

$$\|\sin\Theta(\mathscr{X},\tilde{\mathscr{X}})\| = \|Z\|.$$

代入 (7.30) 即得到估计式(7.27)．□

定理 7.8 ($\sin\theta$ 第一定理的推广)[75]． 在定理 7.7 的假设条件中，如果不再假定 $X=(X_1,X_2)$ 是酉阵，而是仅仅假定X为非奇异阵和 $X_1^H X_1 \geqslant \varepsilon^2 I$，则对于 $\mathbf{C}^{l\times l}$ 上的任一酉不变范数 $\|\ \ \|$，有估计式

$$\|\sin\Theta(\mathscr{X},\tilde{\mathscr{X}})\| \leqslant \frac{\|R\|}{\varepsilon\delta}. \tag{7.31}$$

证明:

首先按照定理 7.7 的证明，在 (7.24) 式两端同时左乘以 $\tilde{X}_2^H$，得到(7.28)以及 $Z=\tilde{X}_2^H X_1$ 的范数估计式(7.30)．

然后，令 $X_1' = X_1(X_1^H X_1)^{-\frac{1}{2}}$． 根据引理 7.1 以及酉不变范数与谱范数的相容性，得出

$$\|\sin\Theta(\mathscr{X},\tilde{\mathscr{X}})\| = \|\tilde{X}_2^H X_1'\| = \|\tilde{X}_2^H X_1(X_1^H X_1)^{-\frac{1}{2}}\|$$
$$\leqslant \|(X_1^H X_1)^{-\frac{1}{2}}\|_2\|\tilde{X}_2^H X_1\| \leqslant \frac{\|Z\|}{\varepsilon}.$$

再将(7.30)代入上式的右端，即得到估计式(7.31)．□

定理 7.9 ($\sin\theta$ 第二定理)[75]． 在定理 7.7 的假设条件中，如果把条件(7.26)改为

$$\delta \equiv \min_{i,j}\{|\lambda_i - \tilde{\lambda}_j| : \lambda_i \in \lambda(A_{11}), \tilde{\lambda}_j \in \lambda(\tilde{A}_{22})\} > 0, \tag{7.32}$$

则

$$\|\sin\Theta(\mathscr{X},\tilde{\mathscr{X}})\|_F \leqslant \frac{\|R\|_F}{\delta}. \tag{7.33}$$

证明:

首先将 A_{11} 与 $\tilde{A}_{22}$ 分解:

$$A_{11} = U_1^H \Lambda_1 U_1, \quad \tilde{A}_{22} = \tilde{U}_2^H \tilde{\Lambda}_2 \tilde{U}_2,$$

其中 U_1 与 $\tilde{U}_2$ 为酉阵，$\Lambda_1 = \text{diag}(\lambda_1, \cdots, \lambda_l)$，$\tilde{\Lambda}_2 = \text{diag}(\tilde{\lambda}_{l+1}, \cdots, \tilde{\lambda}_n)$. 代入方程(7.29)，并令

$$\tilde{U}_2 Z U_1^H = Y, \quad \tilde{U}_2 \tilde{X}_2^H R U_1^H = C,$$

则方程(7.29)变为

$$\tilde{\Lambda}_2 Y - Y\Lambda_1 = C, \quad Y \in \mathbf{C}^{(n-l)\times l} \tag{7.34}$$

如同第一章§5中5.2所做的那样，将 Y 与 C 的元素分别排列成 $l(n-l)$ 维列向量 $\boldsymbol{y}$ 与 $\mathbf{c}$，则方程(7.34)可改写为（见第一章(5.3)式）

$$(I^{(l)} \otimes \tilde{\Lambda}_2 - \Lambda_1 \otimes I^{(n-l)})\boldsymbol{y} = \mathbf{c}.$$

取范数$\| \quad \|_2$，并利用(7.32)，从上式得到

$$\delta \|\boldsymbol{y}\|_2 \leqslant \|(I^{(l)} \otimes \tilde{\Lambda}_2 - \Lambda_1 \otimes I^{(n-l)})\boldsymbol{y}\|_2 = \|\mathbf{c}\|_2. \tag{7.35}$$

再注意到

$$\|\boldsymbol{y}\|_2 = \|Y\|_F = \|Z\|_F = \|\sin\Theta(\mathscr{X}, \tilde{\mathscr{X}})\|_F$$

和

$$\|\mathbf{c}\|_2 = \|C\|_F \leqslant \|R\|_F,$$

代入(7.35)式，便得到估计式(7.33). □

仿定理 7.8 的证明，可证

定理 7.10 ($\sin\theta$ 第二定理的推广)[75]. 在定理 7.9 的假设条件中，如果不再假定 $X = (X_1, X_2)$ 是酉阵，而是仅仅假定 X 是非奇异阵，和 $X_1^H X_1 \geqslant \varepsilon^2 I$，则有

$$\|\sin\Theta(\mathscr{X}, \tilde{\mathscr{X}})\|_F \leqslant \frac{\|R\|_F}{\varepsilon\delta}. \tag{7.36}$$

定理 7.6 的证明（应用定理 7.9）:

首先分别取定理 7.6 中 Z_1 与 Q_1 的正交补 Z_2 与 Q_2，使得 $Z = (Z_1, Z_2)$ 与 $Q = (Q_1, Q_2)$ 为 n 阶酉阵. 据定理假设，存在 A_1 与 A_2，满足

$$Z^H A Z = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad A_1 \in \mathbf{C}^{l\times l}.$$

现构造

$$\widetilde{A} = Q\begin{pmatrix} Q_1^H A Q_1 & 0 \\ 0 & A_2 \end{pmatrix} Q^H,$$

即

$$Q^H \widetilde{A} Q = \begin{pmatrix} Q_1^H A Q_1 & 0 \\ 0 & A_2 \end{pmatrix}.$$

然后分别把这里的 $\widetilde{A}$ 与 A 当作定理 7.9 中的 A 与 $\widetilde{A}$，再分别把这里的 $Q_1^H A Q_1$ 与 A_2 分别当作定理 7.9 中的 A_{11} 与 $\widetilde{A}_{22}$，同时，这里的 Q_1 与 Z_1 分别是定理 7.9 中的 X_1 与 $\widetilde{X}_1$. 于是，定理 7.9 的条件全部满足，因此可以利用估计式（7.33），立即得出不等式(7.17). □

7.3 与近似特征空间有关的其它估计

定义 7.4 和定理 7.4 一定理 7.6 表明：如果 $R(X_1)$ 是 Hermite 阵 A 的一个特征空间，并且 $X_1^H X_1 = I^{(l)}$，则 Rayleigh 商 $A_1 = X_1^H A X_1$ 满足 $\lambda(A_1) \subseteq \lambda(A)$；如果 $R(Q_1)$ 是 A 的一个近似特征空间，并且 $Q_1^H Q_1 = I^{(l)}$，则 Rayleigh 商 $H_1 = Q_1^H A Q_1$ 的 l 个特征值必靠近 A 的 l 个特征值 $\lambda_{1'}, \cdots, \lambda_{l'}$；此外，由 A 的属于 $\lambda_{1'}, \cdots, \lambda_{l'}$ 的特征向量所张成的特征空间 $R(Z_1)$ 与 $R(Q_1)$ 之间的差距，以及 H_1 的 l 个特征值与 $\lambda_{1'}, \cdots, \lambda_{l'}$ 的差距，都可以借助于剩余

$$R = AQ_1 - Q_1 H_1$$

的范数给出上界.

如果 $Q_1 \in \mathbf{C}^{n \times l}$ 满足 $Q_1^H Q_1 = I^{(l)}$，Hermite 阵 $H_1 \in \mathbf{C}^{l \times l}$（不再限于 A 的 Rayleigh 商 $Q_1^H A Q_1$）使得 $R = AQ_1 - Q_1 H_1$ 的范数很小，问：H_1 的 l 个特征值是否必靠近 A 的 l 个特征值？$R(Q_1)$ 是否必靠近 A 的某一个特征空间？其间的差距是否可以用 R 的范数界定？这正是本节要讨论的问题(下面的定理 7.11—7.13，参考 [136]第 11 章).

首先引证关于保范扩张的一条定理.

定理 7.11 (Kahan, 1967). 设 $R=\begin{pmatrix}H\\B\end{pmatrix}$, 其中$H$是 Hermite 阵. 则存在 Hermite 阵 W, 使得"扩张"后的矩阵

$$A=\begin{pmatrix}H & B^H\\ B & W\end{pmatrix}$$

满足 $\|A\|_2=\|R\|_2$.

证明:

对于待定的 W, 记

$$\tilde{R}=\begin{pmatrix}B^H\\ W\end{pmatrix},\quad A=(R,\tilde{R}),\quad \rho=\|R\|_2.$$

由

$$A^2=\begin{pmatrix}R^HR & R^H\tilde{R}\\ \tilde{R}^HR & \tilde{R}^H\tilde{R}\end{pmatrix}$$

可以看出,不管W如何选取,总有

$$\rho^2\leqslant\|A\|_2^2;$$

而定理要求选取 W, 使得 $\rho^2I\geqslant A^2$.

首先证明,对于任意的 $\sigma>\rho$,可选取, $W=W(\sigma)$, 使得 $\sigma^2I-A^2>0$. 证明如下. 利用块三角分解, 可以得出

$$\sigma^2I-A^2=\begin{pmatrix}I & 0\\ L & I\end{pmatrix}\begin{pmatrix}\sigma^2I-R^HR & 0\\ 0 & U(\sigma)\end{pmatrix}\begin{pmatrix}I & L^H\\ 0 & I\end{pmatrix}\tag{7.37}$$

和

$$\sigma^2I-RR^H=\begin{pmatrix}I & 0\\ K & I\end{pmatrix}\begin{pmatrix}\sigma^2I-H^2 & 0\\ 0 & V(\sigma)\end{pmatrix}\begin{pmatrix}I & K^H\\ 0 & I\end{pmatrix},$$

其中

$$U(\sigma)=\sigma^2[I-\tilde{R}^H(\sigma^2I-RR^H)^{-1}\tilde{R}],\tag{7.38}$$

$$V(\sigma)=\sigma^2[I-B(\sigma^2I-H^2)^{-1}B^H]\tag{7.39}$$

和

$$L=-\tilde{R}^HR(\sigma^2I-R^HR)^{-1},\quad K=-BH(\sigma^2I-H^2)^{-1};$$

并且由 $\sigma>\rho=\|R\|_2$ 可知

$$\sigma^2 I - R^H R > 0,\ \sigma^2 I - RR^H > 0,\ \sigma^2 I - H^2 > 0,$$

因此必有 $V(\sigma) > 0$. 现取

$$W \equiv W(\sigma) = -BH(\sigma^2 I - H^2)^{-1}B^H, \tag{7.40}$$

可以验证（本节习题 2），对于如此选取的 W，有 $W^H = W$ 和 $U(\sigma) = V(\sigma) > 0$. 从而 $\sigma^2 I - A^2 > 0$.

然后，令 $\sigma \to \rho + 0$. 注意到(7.40)所示的 $W(\sigma)$ 是复变量 σ 的有理函数，因而是 σ 的亚纯函数；它的奇点只有极点，也就是那些使得 $\|W(\sigma)\|_2$ 无界的点 $\sigma \in \mathbb{C}$. 但是对于实轴上 $\sigma > \rho$ 的部分，恒有 $\|W\|_2 \leqslant \|A\|_2 < \sigma$，因此 $W(\sigma)$ 在 $\sigma = \rho$ 必解析，从而存在

$$\lim_{\sigma \to \rho + 0} W(\sigma) = W(\rho).$$

所以对于如此选取的 $W = W(\rho)$，满足 $\rho^2 I \geqslant A^2$. □

定理 7.12 (Kahan, 1967). 设 $A \in \mathbb{C}^{n\times n}$ 与 $H_1 \in \mathbb{C}^{l\times l}$ ($1 \leqslant l \leqslant n-1$) 均为 Hermite 阵，$\lambda(A) = \{\lambda_i\}_{i=1}^n$，$\lambda(H_1) = \{\mu_i\}_{i=1}^l$. 又设 $Q_1 \in \mathbb{C}^{n\times l}$，$Q_1^H Q_1 = I$. 令 $R = AQ_1 - Q_1 H_1$. 则存在 A 的 l 个特征值 $\lambda_{1'}, \cdots, \lambda_{l'}$，使得

$$|\lambda_{i'} - \mu_i| \leqslant \|R\|_2,\ i = 1, \cdots, l. \tag{7.41}$$

证明:

无妨假设 $Q = \begin{pmatrix} I^{(l)} \\ 0 \end{pmatrix}$. 记

$$A = \begin{pmatrix} A_{11} & A_{21}^H \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} H_1 & 0 \\ 0 & A_{22} - W \end{pmatrix} + \begin{pmatrix} A_{11} - H_1 & A_{21}^H \\ A_{21} & W \end{pmatrix},$$

其中 W 是根据定理 7.11 选取的 $n-l$ 阶 Hermite 阵，使得

$$\left\| \begin{pmatrix} A_{11} - H_1 & A_{21}^H \\ A_{21} & W \end{pmatrix} \right\|_2 = \left\| \begin{pmatrix} A_{11} - H_1 \\ A_{21} \end{pmatrix} \right\|_2. \tag{7.42}$$

应用 Weyl-Лидский 定理，存在 $\lambda_{i'} \in \lambda(A)$，$i = 1, \cdots, l$，使得

$$|\lambda_{i'} - \mu_i| \leqslant \left\| \begin{pmatrix} A_{11} - H_1 & A_{21}^H \\ A_{21} & W \end{pmatrix} \right\|_2; \tag{7.43}$$

另一方面，有

$$R = A\begin{pmatrix} I \\ 0 \end{pmatrix} - \begin{pmatrix} I \\ 0 \end{pmatrix} H_1 = \begin{pmatrix} A_{11} - H_1 \\ A_{21} \end{pmatrix}. \tag{7.44}$$

由(7.42)—(7.44)立即导出(7.41). □

定理 7.13 (Kahan, 1967). 设 $A \in \mathbf{C}^{n\times n}$ 与 $H_1 \in \mathbf{C}^{l\times l}$ ($1 \leqslant l \leqslant n-1$) 均为 Hermite 阵. $\lambda(A) = \{\lambda_i\}_{i=1}^n$, $\lambda(H_1) = \{\mu_i\}_{i=1}^l$. 又设 $Q_1 \in \mathbf{C}^{n\times l}$ 满足

$$Q_1^H Q_1 \geqslant \varepsilon^2 I, \quad \varepsilon > 0. \tag{7.45}$$

令 $R = AQ_1 - Q_1H_1$. 则存在 $\lambda_{1'}, \cdots, \lambda_{l'} \in \lambda(A)$, 使得

$$|\lambda_{i'} - \mu_i| \leqslant \frac{\sqrt{2}\,\|R\|_2}{\varepsilon}. \tag{7.46}$$

证明

无妨假设 $Q_1 = \begin{pmatrix} \Sigma \\ 0 \end{pmatrix}$, 其中 $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_l), \sigma_1 \geqslant \cdots \geqslant \sigma_l$ 是 Q_1 的奇异值, $\sigma_l > 0$. 记

$$A = \begin{pmatrix} A_{11} & A_{21}^H \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} H_1 & 0 \\ 0 & A_{22} - W \end{pmatrix} + \begin{pmatrix} A_{11} - H_1 & A_{21}^H \\ A_{21} & W \end{pmatrix},$$

其中 W 是根据定理 7.11 选取的 $n-1$ 阶 Hermite 阵,满足(7.42)式. 应用 Weyl-Лидский 定理,存在 $\lambda_{1'}, \cdots, \lambda_{l'} \in \lambda(A)$, 使得

$$\left.\begin{aligned} |\lambda_{i'} - \mu_i| &\leqslant \left\| \begin{pmatrix} A_{11} - H & A_{21}^H \\ A_{21} & W \end{pmatrix} \right\|_2 = \|\hat{R}\|_2, \\ \hat{R} &= \begin{pmatrix} A_{11} - H \\ A_{21} \end{pmatrix}. \end{aligned}\right\} \tag{7.47}$$

另一方面,有

$$R = A\begin{pmatrix} \Sigma \\ 0 \end{pmatrix} - \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} H_1 = \begin{pmatrix} A_{11}\Sigma - \Sigma H_1 \\ A_{21}\Sigma \end{pmatrix}. \tag{7.48}$$

以下比较 $\|\hat{R}\|_2$ 与 $\|R\|_2$.

首先,有

$$\|A_{21}\|_2 \leqslant \|A_{21}\Sigma\|_2 / \sigma_l. \tag{7.49}$$

其次考察

$$A_{11}\Sigma - \Sigma H_1 = (A_{11} - H_1)\Sigma + (H_1\Sigma - \Sigma H_1).$$

设 $A_{11}-H_1$ 的最大模特征值为 λ，相应的单位特征向量为 $\boldsymbol{x}$，即

$$\|A_{11}-H_1\|_2=|\lambda|,\ (A_{11}-H_1)\boldsymbol{x}=\lambda\boldsymbol{x},\ \|\boldsymbol{x}\|_2=1. \tag{7.50}$$

于是有

$$\begin{aligned}\boldsymbol{x}^H(A_{11}\Sigma-\Sigma H_1)\boldsymbol{x}&=\boldsymbol{x}^H(A_{11}-H_1)\Sigma\boldsymbol{x}+\boldsymbol{x}^H(H_1\Sigma-\Sigma H_1)\boldsymbol{x}\\&=\boldsymbol{x}^H(A_{11}-H_1)\Sigma\boldsymbol{x}+i\tau(\because H_1\Sigma-\Sigma H_1\ \text{是斜 Hermite 阵})\\&=\lambda\boldsymbol{x}^H\Sigma\boldsymbol{x}+i\tau\quad(\tau\in\mathbb{R}).\end{aligned} \tag{7.51}$$

从而

$$\begin{aligned}\|A_{11}\Sigma-\Sigma H_1\|_2&=\max_{\substack{\|\boldsymbol{u}\|_2=1\\\|\boldsymbol{v}\|_2=1}}|\boldsymbol{u}^H(A_{11}\Sigma-\Sigma H_1)\boldsymbol{v}|\ (\text{见 § 3 习题 2})\\&\geqslant\max_{\|\boldsymbol{u}\|_2=1}|\boldsymbol{u}^H(A_{11}\Sigma-\Sigma H_1)\boldsymbol{u}|\\&\geqslant|\boldsymbol{x}^H(A_{11}\Sigma-\Sigma H_1)\boldsymbol{x}|\\&\geqslant|\lambda||\boldsymbol{x}^H\Sigma\boldsymbol{x}|\qquad(\text{利用 (7.51)})\\&=\|A_{11}-H_1\|_2|\boldsymbol{x}^H\Sigma\boldsymbol{x}|\quad(\text{利用(7.50)})\\&\geqslant\|A_{11}-H_1\|_2\sigma_l.\end{aligned} \tag{7.52}$$

因此，从(7.47)—(7.49)和(7.52)得到

$$\begin{aligned}\|\hat{R}\|_2^2&=\|\hat{R}^H\hat{R}\|_2=\|(A_{11}-H_1)^2+A_{21}^HA_{21}\|_2\\&\leqslant\|A_{11}-H_1\|_2^2+\|A_{21}\|_2^2\\&\leqslant\frac{\|A_{11}\Sigma-\Sigma H\|_2^2+\|A_{21}\Sigma\|_2^2}{\sigma_l^2}\leqslant\frac{2\|R\|_2^2}{\sigma_l^2}.\end{aligned}$$

代入(7.47)便导出估计式(7.46). □

定理 7.14. 在定理 7.12 的假设条件下，如果 A 属于 $\lambda_{1'},\cdots,\lambda_{l'}$ 的特征向量所张成的特征空间为 $R(Z_1)$，$Z_1^HZ_1=I^{(l)}$，并且如果对于 $\mathscr{I}=\{1,\cdots,n\}\backslash\{1',\cdots,l'\}$ 有

$$\delta\equiv\min\{|\lambda_i-\mu_j|:i\in\mathscr{I},j=1,\cdots,l\}>0, \tag{7.53}$$

则

$$\|\sin\Theta(R(Z_1),\ R(Q_1))\|_F\leqslant\frac{\|R\|_F}{\delta}. \tag{7.54}$$

证明:

首先，分别取 Z_2 与 Q_2，使 $Z=(Z_1,Z_2)$ 与 $Q=(Q_1,Q_2)$

皆为酉阵．据题设，存在 A_1 与 A_2，满足

$$Z^H A Z = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad A_1 \in \mathbf{C}^{l\times l}.$$

现构造

$$\tilde{A} = Q\begin{pmatrix} H_1 & 0 \\ 0 & A_2 \end{pmatrix} Q^H,$$

即

$$Q^H \tilde{A} Q = \begin{pmatrix} H_1 & 0 \\ 0 & A_2 \end{pmatrix}.$$

然后分别把这里的 $\tilde{A}$ 与 A 当作定理 7.9 中的 A 与 $\tilde{A}$，再分别把这里的 H_1, A_2, Q_1 与 Z_1 当作定理 7.9 中的 A_{11}, $\tilde{A}_{22}$, X_1 与 $\tilde{X}_1$. 容易验证，定理 7.9 的条件全部被满足，因此可利用 (7.33) 立即得出 (7.54). □

定理 7.15. 在定理 7.13 的假设条件下，如果 A 属于 $\lambda_{1'},\cdots,\lambda_{l'}$ 的特征向量所张成的特征空间为 $R(Z_1)$，其中 $Z_1^H Z_1 = I^{(l)}$，并且如果(7.53)式成立，则

$$\|\sin\Theta(R(Z_1), R(Q_1))\|_F \leqslant \frac{\|R\|_F}{\varepsilon\delta}. \tag{7.55}$$

证明:

首先取 Z_2，使 $Z=(Z_1,Z_2)$ 为酉阵；再取 Q_2，使 $Q=(Q_1, Q_2)$ 为非奇异阵．据题设，存在 A_1 与 A_2，满足

$$Z^H A Z = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad A_1 \in \mathbf{C}^{l\times l}.$$

现构造

$$\tilde{A} = Q^{-H}\begin{pmatrix} H_1 & 0 \\ 0 & A_2 \end{pmatrix} Q^{-1},$$

即

$$Q^H \tilde{A} Q = \begin{pmatrix} H_1 & 0 \\ 0 & A_2 \end{pmatrix}.$$

然后把这里的 $\tilde{A}$ 与 A 分别当作定理 7.10 中的 A 与 $\tilde{A}$，再分别把这

里的 H_1, A_2, Q_1 与 Z_1 分别当作定理 7.10 中的 A_{11}, $\tilde{A}_{22}$, X_1 与 $\tilde{X}_1$. 容易验证,定理 7.10 的条件全部被满足,因此可利用(7.36)立即得出(7.55). □

顺便指出两点:1).不等式(7.46)右端中的 $\sqrt{2}$ 能不能用 1 代替,是一个尚未解决的问题. 2). 对于一般的矩阵,与 7.3 这一段内容有关的研究,可参看文献[116].

习题

1.设有矩阵方程

$$BX - XA = C, \quad X \in \mathbf{C}^{m\times n},$$

其中 $A \in \mathbf{C}^{n\times n}$ 与 $B \in \mathbf{C}^{m\times m}$ 为 Hermite 阵, $C \in \mathbf{C}^{m\times n}$. 试证. 如果

$$|\lambda - \mu| \geqslant \delta > 0, \ \forall \lambda \in \lambda(A), \ \mu \in \lambda(B),$$

则有估计式

$$\|X\|_F \leqslant \frac{\|C\|_F}{\delta}$$

和

$$\|X\|_2 \leqslant \frac{\sqrt{\operatorname{rank} C}\,\|C\|_2}{\delta}.$$

2.验证: 按 (7.38) 选取的 W, 满足 $W^H = W$ 和 $U(\sigma) = V(\sigma) > 0$. 其中 $U(\sigma)$ 与 $V(\sigma)$ 分别见 (7.38) 式和(7.39)式.

3.设

$$A = \begin{pmatrix} H_1 & C^H \\ C & H_2 \end{pmatrix} \in \mathbf{C}^{n\times n}, \ \lambda(A) = \{\lambda_i\}_{i=1}^n,$$

其中 $H_1^H = H_1 \in \mathbf{C}^{l\times l}$, $H_2^H = H_2$, $\lambda(H_1) = \{\mu_i\}_{i=1}^l$. 试证: 存在 A 的 l 个特征值 $\lambda_{1'}, \cdots, \lambda_{l'}$, 使得

$$|\lambda_{i'} - \mu_i| \leqslant \|C\|_2, \ i = 1, 2, \cdots, l.$$

§ 8 不变子空间的扰动界限

8.1 不变子空间

定义 8.1. 设 $A\in\mathbb{C}^{n\times n}$. 如果 $\mathbb{C}^n$ 的子空间 $\mathscr{X}$ 满足

$$A\mathscr{X}\subseteq\mathscr{X},$$

则称 $\mathscr{X}$ 为 A 的一个不变子空间.

例如，任一矩阵 A 的一个特征向量所张成的 1 维子空间是 A 的不变子空间；Hermite 阵 A 的特征空间是 A 的不变子空间.

设 $A\in\mathbb{C}^{n\times n}$ 的 Jordan 分解为

$$X^{-1}AX=J,\ 即\ AX=XJ,$$

其中 $X=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)$, Jordan 标准形

$$J=\operatorname{diag}(J_1(\lambda_1),\cdots,J_{s_1}(\lambda_1),J_{s_1+1}(\lambda_2),\cdots,J_{s_1+s_2}(\lambda_2),\\ \cdots,J_{s_{r-1}+1}(\lambda_r),\cdots,J_{s_{r-1}+s_r}(\lambda_r)),$$

每个 $J_{s_{i-1}+j}(\lambda_i)=\begin{pmatrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}$ $(s_0=0,1\leqslant j\leqslant s_i,1\leqslant i\leqslant r)$

是 Jordan 块，并且 $\lambda_p\neq\lambda_q(p\neq q)$. 设 $J_1(\lambda_1)\in\mathbb{C}^{m\times m}$, 于是有链式关系

$$A\boldsymbol{x}_1=\lambda_1\boldsymbol{x}_1,\ A\boldsymbol{x}_2=\lambda\boldsymbol{x}_2+\boldsymbol{x}_1,\ A\boldsymbol{x}_3=\lambda_1\boldsymbol{x}_3+\boldsymbol{x}_2,\\ \cdots,\ A\boldsymbol{x}_m=\lambda\boldsymbol{x}_m+\boldsymbol{x}_{m-1},$$

即

$$\left.\begin{aligned}(A-\lambda_1I)\boldsymbol{x}_1&=0,\\(A-\lambda_1I)\boldsymbol{x}_2&=\boldsymbol{x}_1,\\(A-\lambda_1I)\boldsymbol{x}_3&=\boldsymbol{x}_2,\\&\cdots\cdots\\(A-\lambda_1I)\boldsymbol{x}_m&=\boldsymbol{x}_{m-1}.\end{aligned}\right\}\qquad(8.1)$$

在(8.1)中，第一个关系式表示 $\boldsymbol{x}_1$ 是 A 的特征向量，而其余的关系式表示 $\boldsymbol{x}_k(2\leqslant k\leqslant m)$ 满足

$$(A-\lambda_1I)^{k-1}\boldsymbol{x}_k\neq 0,\ (A-\lambda_1I)^k\boldsymbol{x}_k=0.$$

凡满足

$$(A-\lambda I)^{p-1}\boldsymbol{x}\neq 0,\quad (A-\lambda I)^{p}\boldsymbol{x}=0$$

的向量 $\boldsymbol{x}$，通常叫做 A 属于 λ 的 p 级主向量（或 p 级根向量）．(8.1)式表明，A 的 1 级主向量(即特征向量) $\boldsymbol{x}_1$ 所张成的子空间是 A 的不变子空间，其它的任一主向量 $\boldsymbol{x}_k(2\leqslant k\leqslant m)$ 所张成的子空间不是 A 的不变子空间；但另一方面，由 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m$ 的全体所张成的子空间是 A 的不变子空间．

记 $X_1=(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_m)$．(8.1)式表示

$$AX_1=X_1J_1(\lambda_1).$$

显然 $R(X_1)$ 是 A 的一个不变子空间．这同时建议我们对于不变子空间给出下述定义．

定义 8.2[6]．设 $A\in\mathbf{C}^{n\times n}$，$A_1\in\mathbf{C}^{l\times l}$ $(1\leqslant l\leqslant n-1)$．如果存在 $X_1\in\mathbf{C}^{n\times l}$，适合

$$\text{(i) } \operatorname{rank}X_1=l,\quad \text{(ii) } AX_1=X_1A_1,$$

则 $R(X_1)$ 叫做 A 的不变子空间．

容易看出定义 8.1 和定义 8.2 等价．

在 §7 中的 7.2，我们讨论了 Hermite 阵 A 的特征空间扰动的 Davis-Kahan $\sin\theta$ 定理，在那里曾对 A 以及扰动后的 Hermite 阵 $\tilde{A}$ 的特征值的分布加了一些限制，然后求出 A 与 $\tilde{A}$ 的对应特征空间距离的上界．本节介绍不变子空间扰动的 Stewart 理论，其主要之点在于论证靠近 A 的矩阵 $A+E$ 的对应不变子空间的存在性，并给出两个对应子空间的距离的上界．不过在讨论时，对于 A 的特征值分布加了如下的限制，即：如果研究 A 的不变子空间 $\mathscr{X}$ 的扰动，则事先假定 A 对应于 $\mathscr{X}$ 的特征值与 A 的其余的特征值明显分离．下面的例子可以帮助我们理解，为什么要提出这样的限制条件．

例 8.1.

$$A=\begin{pmatrix}1-\varepsilon & 0 & 0\\ 0 & 1+2\varepsilon & 0\\ 0 & 0 & 2\end{pmatrix},0<\varepsilon\ll 1.$$

显然有

$$AX = X\begin{pmatrix} 1-\varepsilon & 0 & 0 \\ 0 & 1+2\varepsilon & 0 \\ 0 & 0 & 2 \end{pmatrix}, X = (\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3) \equiv I^{(3)}.$$

设

$$\widetilde{A} = A + \varepsilon B,$$

$$B = \begin{pmatrix} \dfrac{9+25\varepsilon+25\varepsilon^2}{5(1+5\varepsilon^2)} & \dfrac{-2}{5\sqrt{1+5\varepsilon^2}} & \dfrac{-5+4\varepsilon}{\sqrt{5}\,(1+5\varepsilon^2)} \\ * & -\dfrac{9}{5} & \dfrac{-2\varepsilon}{\sqrt{5\,(1+5\varepsilon^2)}} \\ * & * & \dfrac{4\varepsilon^2}{1+5\varepsilon^2} \end{pmatrix} = B^T.$$

计算可知

$$\widetilde{A}\widetilde{X} = \widetilde{X}\begin{pmatrix} 1 & & \\ & 1+\varepsilon & \\ & & 2 \end{pmatrix}, \ \widetilde{X} = (\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2, \tilde{\boldsymbol{x}}_3),$$

其中

$$\tilde{\boldsymbol{x}}_1 = \left(\frac{1}{\sqrt{5(1+5\varepsilon^2)}}, \quad \frac{2}{\sqrt{5}}, \ \frac{\varepsilon}{\sqrt{1+5\varepsilon^2}}\right)^T,$$

$$\tilde{\boldsymbol{x}}_2 = \left(\frac{2}{\sqrt{5(1+5\varepsilon^2)}}, \ -\frac{1}{\sqrt{5}}, \ \frac{2\varepsilon}{\sqrt{1+5\varepsilon^2}}\right)^T,$$

$$\tilde{\boldsymbol{x}}_3 = \left(\frac{-5\varepsilon}{\sqrt{5(1+5\varepsilon^2)}}, \quad 0, \quad \frac{1}{\sqrt{1+5\varepsilon^2}}\right)^T,$$

满足

$$\widetilde{X}\widetilde{X}^T = I^{(3)}.$$

容易算出

$$\sin\theta(R(\boldsymbol{e}_1), R(\tilde{\boldsymbol{x}}_1)) = \sqrt{\frac{4+25\varepsilon^2}{5+25\varepsilon^2}} \approx \frac{2}{\sqrt{5}}$$

和

$$\sin\theta(R(\boldsymbol{e}_2), R(\tilde{\boldsymbol{x}}_2)) = \frac{2}{\sqrt{5}}.$$

这些结果表明，A经过一个小的扰动之后，$R(\boldsymbol{e}_1)$ 与 $R(\boldsymbol{e}_2)$ 发生了很大的改变. 可是另一方面，A的特征空间 $\mathscr{X} = R(\boldsymbol{e}_1,\boldsymbol{e}_2)$ 和 $R(\boldsymbol{e}_3)$ 倒是相当稳定的，因为如果记 $\widetilde{\mathscr{X}} = R(\tilde{\boldsymbol{x}}_1,\tilde{\boldsymbol{x}}_2)$ 和

$$\widetilde{X}_{11} = \begin{pmatrix} \frac{1}{\sqrt{5(1+5\varepsilon^2)}} & \frac{2}{\sqrt{5(1+5\varepsilon^2)}} \\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix},$$

则可算出

$$\begin{aligned} \|\sin\Theta(\mathscr{X},\widetilde{\mathscr{X}})\|_2 &= \|I - \widetilde{X}_{11}\widetilde{X}_{11}^H\|_2^{\frac{1}{2}} \\ &= \left\|\begin{pmatrix} \frac{5\varepsilon^2}{1+5\varepsilon^2} & 0 \\ 0 & 0 \end{pmatrix}\right\|_2^{\frac{1}{2}} = \sqrt{\frac{5}{1+5\varepsilon^2}}\cdot\varepsilon \approx \sqrt{5}\,\varepsilon, \end{aligned}$$

同时，有

$$\sin\theta(R(\boldsymbol{e}_3),\ R(\tilde{\boldsymbol{x}}_3)) = \sqrt{\frac{5}{1+5\varepsilon^2}}\cdot\varepsilon \approx \sqrt{5}\,\varepsilon.$$

究其原因，可以从几何上进行如下的考察：A表示 $\mathbb{R}^3$ 中以原点为中心、轴线落在三个坐标轴上、半轴长度分别为 $1-\varepsilon$, $1+2\varepsilon$ 和 2 的椭球面(这时，把A和 $\mathbb{R}^3$ 中的椭球面

$$\mathscr{S}_A = \{A\boldsymbol{x}:\boldsymbol{x}\in\mathbb{R}^3,\ \|\boldsymbol{x}\|_2 = 1\}$$

等同起来)；经过一个 ε 量级的扰动后，椭球面A变成了椭球面 $\widetilde{A}$. 由于A的第一、第二两个半轴的长度很接近，同时它们与第三个半轴的长度相差较大，因此经过 ε 扰动后，虽然半轴的长度变化不大(由 $1-\varepsilon$, $1+2\varepsilon$ 和 2 分别变成了 1，$1+\varepsilon$ 和 2)，但第一、第二两个半轴的方位却可以发生较大的改变；而另一方面，椭球面A的第一、第二两个半轴所在平面的方位，以及第三个半轴的方位，变化并不大(如图 3-3 所示).

所以，如果A的一个不变子空间 $\mathscr{X}$ 所对应的特征值与A的其它特征值明显分离，那么就有可能对于 $\mathscr{X}$ 的扰动界限求出一些有用的信息.

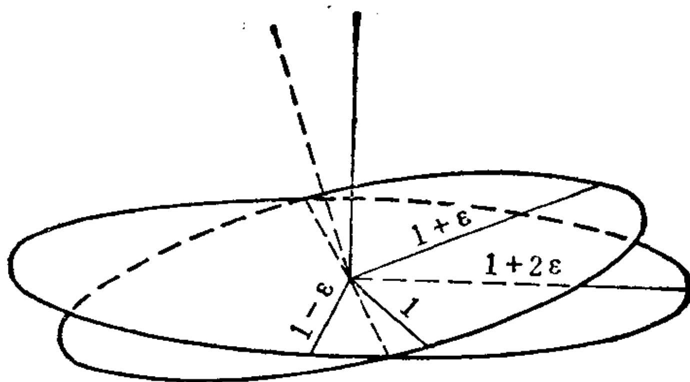


图 3-3

附带说明，我们在本节内将使用 $\bigcup\limits_{m,n=1}^{N} \mathbb{C}^{m\times n}$ 上相容的范数 $\|\ \ \|$，其中 N 是一个足够大的自然数．所谓 $\|\ \ \|$ 是 $\bigcup\limits_{m,n=1}^{N} \mathbb{C}^{m\times n}$ 上相容的范数，其定义是：如果把 $\|\ \ \|$ 限制到每一个 $\mathbb{C}^{m\times n}(m,n\leqslant N)$ 上，则 $\|\ \ \|$ 是 $\mathbb{C}^{m\times n}$ 上的范数，并且只要乘积 AB 有意义，则

$$\|AB\| \leqslant \|A\|\|B\|.$$

例如，$\|\ \ \|_1$，$\|\ \ \|_2$，$\|\ \ \|_\infty$，$\|\ \ \|_F$ 都是 $\bigcup\limits_{m,n=1}^{N} \mathbb{C}^{m\times n}$ 上相容的范数．又如，若 $\|\ \ \|$ 是 $\mathbb{C}^{N\times N}$ 上的酉不变范数，它由 $\mathbb{R}^N$ 上的 SG 函数 $\Phi_N(\xi_1,\cdots,\xi_N)$ 生成，则可如下定义一个在 $\bigcup\limits_{m,n=1}^{N} \mathbb{C}^{m\times n}$ 上相容的酉不变范数 $\|\ \ \|$：对于任一 $A\in\mathbb{C}^{m\times n}(m,n\leqslant N)$，设 A 的奇异值分解为 $A=U\Sigma V^H$，其中 U 与 V 为酉阵，$\Sigma=\mathrm{diag}\,(\sigma_1,\sigma_2,\cdots)$，$\sigma_1\geqslant\cdots\geqslant\sigma_n$ 是 A 的奇异值，则定义

$$\|A\| = \Phi_N(\sigma_1,\sigma_2,\cdots,\sigma_n,0,\cdots,0).$$

应用到 $\mathbb{C}^{m\times n}$ 上定义的算子，我们将采用算子范数，即

$$\|\mathbf{T}\| \equiv \sup_{\|P\|=1} \|\mathbf{T}P\|,$$

上式右端的范数 $\|\ \ \|$ 是 $\bigcup\limits_{m,n=1}^{N} \mathbb{C}^{m\times n}$ 上相容的矩阵范数．如此定义的算子范数，显然满足

$$\|\mathbf{T}P\| \leqslant \|\mathbf{T}\|\|P\|.$$

8.2 一个非线性方程及其解的估计

设 $A \in \mathbf{C}^{n\times n}$，其分块形式为

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad A_{11} \in \mathbf{C}^{l\times l}, \quad 1 \leqslant l \leqslant n-1.$$

如果 $A_{21} = 0$，则有

$$A\begin{pmatrix} I^{(l)} \\ 0 \end{pmatrix} = \begin{pmatrix} I^{(l)} \\ 0 \end{pmatrix} A_{11},$$

即 $R\left(\begin{pmatrix} I^{(l)} \\ 0 \end{pmatrix}\right)$ 是 A 的一个不变子空间；但是如果 $A_{21} \neq 0$，而是 $\|A_{21}\|$ 很小，问：在什么条件下，A 存在一个靠近 $R\left(\begin{pmatrix} I^{(l)} \\ 0 \end{pmatrix}\right)$ 的不变子空间$\left(\text{即子空间 } R\left(\begin{pmatrix} I^{(l)} \\ 0 \end{pmatrix}\right) \text{ 是 } A \text{ 的一个近似不变子空间}\right)$？这就是 Stewart [152] 和[155]研究不变子空间扰动理论的出发点。

一种自然的考虑是，设想 A 有一个不变子空间 $R\left(\begin{pmatrix} I^{(l)} \\ P \end{pmatrix}\right)$，即存在 $P \in \mathbf{C}^{(n-l)\times l}$ 与 $A_1 \in \mathbf{C}^{l\times l}$，使得

$$A\begin{pmatrix} I \\ P \end{pmatrix} = \begin{pmatrix} I \\ P \end{pmatrix} A_1, \tag{8.2}$$

或者

$$\begin{pmatrix} I & 0 \\ P & I \end{pmatrix}^{-1} A \begin{pmatrix} I & 0 \\ P & I \end{pmatrix} = \begin{pmatrix} A_1 & * \\ 0 & * \end{pmatrix}. \tag{8.3}$$

由(8.3)可知，P 与 A_1 应满足

$$A_{11} + A_{12}P = A_1 \tag{8.4}$$

和

$$PA_{11} - A_{22}P = A_{21} - PA_{12}P. \tag{8.5}$$

如果在某些条件下，能够证明(8.5)存在解 P，然后将这个 P 代入(8.4)算出 A_1，则 P 与 A_1 满足(8.2)。这样得到的 A 的不变子

空间 $R\left(\begin{pmatrix} I \\ P \end{pmatrix}\right)$与子空间 $R\left(\begin{pmatrix} I \\ 0 \end{pmatrix}\right)$的距离为

$$\|\sin\Theta\| = \|[I - (I + PP^H)^{-1}]^{\frac{1}{2}}\|$$
$$= \|[PP^H(I + PP^H)^{-1}]^{\frac{1}{2}}\|, \tag{8.6}$$

其中$\|\ \ \|$是 $\bigcup_{m,n=1}^{N} \mathbf{C}^{m\times n}$ 上相容的酉不变范数，

$$\Theta \equiv \Theta\left(R\left(\begin{pmatrix} I \\ P \end{pmatrix}\right), R\left(\begin{pmatrix} I \\ 0 \end{pmatrix}\right)\right).$$

利用P的奇异值分解 $P = U\Sigma V^H$，其中U与V为酉阵，$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots)$，可算出

$$\sin\Theta = U[\Sigma\Sigma^H(I + \Sigma\Sigma^H)^{-1}]^{\frac{1}{2}}U^H,$$
$$\cos\Theta = U[I + \Sigma\Sigma^H]^{-\frac{1}{2}}U^H;$$

由此得到

$$\mathrm{tg}\,\Theta = U(\Sigma\Sigma^H)^{\frac{1}{2}}U^H.$$

因而，对于 $\bigcup_{m,n=1}^{N} \mathbf{C}^{m\times n}$ 上任一相容的酉不变范数$\|\ \ \|$，有

$$\|P\| = \|\mathrm{tg}\,\Theta\|. \tag{8.7}$$

(8.7)说明了$\|P\|$的几何意义.

如果取$\|\ \ \|$为谱范数，则可由(8.6)及P的奇异值分解得出

$$\|\sin\Theta\|_2 = \|P\|_2/\sqrt{1 + \|P\|_2^2}. \tag{8.8}$$

同时，从(8.6)可知，对 $\bigcup_{m,n=1}^{N} \mathbf{C}^{m\times n}$ 上任一相容的酉不变范数$\|\ \ \|$，恒有

$$\|\sin\Theta\| \leqslant \|P\|. \tag{8.9}$$

因此，问题归结为建立一些条件，使方程(8.5)存在范数很小的解P，并给出$\|P\|$的上界.

把方程(8.5)改记为

$$\mathbf{T}P = A_{21} - \varphi(P), \tag{8.10}$$

其中$\mathbf{T}$是 $\mathbf{C}^{(n-l)\times l}$ 上的一个有界线性算子，φ是 $\mathbf{C}^{(n-l)\times l}$ 到

$\mathbf{C}^{(n-l)\times l}$ 上的一个映射，它们分别由

$$\mathbf{T}P \equiv PA_{11} - A_{22}P \tag{8.11}$$

和

$$\varphi(P) \equiv PA_{12}P \tag{8.12}$$

定义。

下面证明关于一个非线性算子方程求解的一般性结论。

定理 8.1[155]. 设 $\mathbf{T}$ 是 Banach 空间 $\mathscr{B}$ 上的一个有界线性算子，并且 $\mathbf{T}^{-1}$ 有界。设 φ 是 $\mathscr{B}$ 中满足下列条件的连续映射：存在常数 $\eta \geqslant 0$，使得

(i) $\|\varphi(\boldsymbol{x})\| \leqslant \eta\|\boldsymbol{x}\|^2$,

(ii) $\|\varphi(\boldsymbol{x}) - \varphi(\boldsymbol{y})\| \leqslant 2\eta\max(\|\boldsymbol{x}\|, \|\boldsymbol{y}\|)\|\boldsymbol{x} - \boldsymbol{y}\|$.

又设 $g \in \mathscr{B}$，并且

$$\gamma = \|g\|, \quad \delta = \|\mathbf{T}^{-1}\|^{-1}.$$

令 $\boldsymbol{x}_0 = 0$，并定义序列 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots$:

$$\boldsymbol{x}_{i+1} = \mathbf{T}^{-1}(g - \varphi(\boldsymbol{x}_i)) \equiv \boldsymbol{x}_1 - \mathbf{T}^{-1}\varphi(\boldsymbol{x}_i). \tag{8.13}$$

如果

$$\kappa_2 = \gamma\eta/\delta^2 < \frac{1}{4}, \tag{8.14}$$

则 $\boldsymbol{x}_i$ 收敛到方程

$$\mathbf{T}\boldsymbol{x} = g - \varphi(\boldsymbol{x}) \tag{8.15}$$

的一个解 $\boldsymbol{x}$，并且对于

$$\kappa = \frac{2\kappa_2}{1 - 2\kappa_2 + \sqrt{1 - 4\kappa_2}},$$

$\boldsymbol{x}$ 满足

$$\|\boldsymbol{x}\| \leqslant \frac{\gamma}{\delta}(1 + \kappa) < \frac{2\gamma}{\delta}. \tag{8.16}$$

此外，

$$\|\boldsymbol{x} - \boldsymbol{x}_i\| \leqslant \frac{\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\|}{1 - \rho} \leqslant \frac{\rho^{i-k}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|}{1 - \rho},$$

$$k \leqslant i, \tag{8.17}$$

其中 $\rho = 4\kappa_2 < 1$.

证明:

令 $\kappa_1 = 0$, 显然有

$$\|x_1\| = \|T^{-1}g\| \leqslant \|T^{-1}\|\|g\| = \frac{\gamma}{\delta}(1 + \kappa_1).$$

从而

$$\|x_2\| \leqslant \|x_1\| + \|T^{-1}\|\|\varphi(x_1)\| \leqslant \frac{\gamma}{\delta} + \frac{\eta}{\delta}\|x_1\|^2$$

$$\leqslant \frac{\gamma}{\delta}(1 + \kappa_2),$$

其中 κ_2 如(8.14)所示. 一般地,如果令

$$\kappa_{i+1} = \kappa_2(1 + \kappa_i)^2, i = 2,3,\cdots, \tag{8.18}$$

则由

$$\|x_i\| \leqslant \frac{\gamma}{\delta}(1 + \kappa_i) \tag{8.19}$$

可导出

$$\|x_{i+1}\| \leqslant \|x_1\| + \|T^{-1}\|\|\varphi(x_i)\| \leqslant \frac{\gamma}{\delta} + \frac{\eta}{\delta}\cdot\frac{\gamma^2}{\delta^2}(1 + \kappa_i)^2$$

$$\leqslant \frac{\gamma}{\delta}(1 + \kappa_2(1 + \kappa_i)^2) = \frac{\gamma}{\delta}(1 + \kappa_{i+1}).$$

因此,对于 $i = 1,2,\cdots,x_i$ 恒满足(8.19).

利用(8.14)和(8.18),可以归纳地证明序列 $\{\kappa_i\}$ 单调递增,并且 $\kappa_i < 1, i = 1,2,\cdots$, 所以存在

$$\kappa = \lim_{i\to\infty}\kappa_i.$$

由(8.18)可知, $\kappa = \kappa_2(1 + \kappa)^2$, 即

$$\kappa = \frac{2\kappa_2}{1 - 2\kappa_2 + \sqrt{1 - 4\kappa_2}} < 1.$$

因此,整个迭代都位于闭圆

$$\mathscr{D} = \left\{x \in \mathscr{B} : \|x\| \leqslant \frac{\gamma}{\delta}(1 + \kappa)\right\}$$

之中.

再由

$$\|\boldsymbol{x}_{i+1}-\boldsymbol{x}_i\| \leqslant \|\mathbf{T}^{-1}\|\|\varphi(\boldsymbol{x}_i)-\varphi(\boldsymbol{x}_{i-1})\|$$

$$\leqslant \frac{1}{\delta}\cdot 2\eta\cdot\frac{\gamma}{\delta}(1+\kappa)\|\boldsymbol{x}_i-\boldsymbol{x}_{i-1}\|$$

$$=2(1+\kappa)\kappa_2\|\boldsymbol{x}_i-\boldsymbol{x}_{i-1}\|<4\kappa_2\|\boldsymbol{x}_i-\boldsymbol{x}_{i-1}\|$$

$$=\rho\|\boldsymbol{x}_i-\boldsymbol{x}_{i-1}\|\leqslant\rho^i\|\mathbf{T}^{-1}g\|$$

及 $\rho=4\kappa_2<1$ 立即得知，存在 $\boldsymbol{x}=\lim\limits_{i\to\infty}\boldsymbol{x}_i$，并且

$$\|\boldsymbol{x}\|\leqslant\frac{\gamma}{\delta}(1+\kappa)<\frac{2\gamma}{\delta}.$$

此外，可得估计式

$$\|\boldsymbol{x}-\boldsymbol{x}_i\|=\left\|\sum_{j=i}^{\infty}(\boldsymbol{x}_{j+1}-\boldsymbol{x}_j)\right\|\leqslant\sum_{j=i}^{\infty}\rho^{j-i}\|\boldsymbol{x}_{i+1}-\boldsymbol{x}_i\|$$

$$\leqslant\frac{\|\boldsymbol{x}_{i+1}-\boldsymbol{x}_i\|}{1-\rho}\leqslant\frac{\rho^{i-k}\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_k\|}{1-\rho},\ k\leqslant i.\ \square$$

应用定理 8.1 可知，对于 $A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}$ 和 (8.11) 所定义的 $\mathbf{T}$，令 $\delta=\|\mathbf{T}^{-1}\|^{-1}$，$\gamma=\|A_{21}\|$，$\eta=\|A_{12}\|$. 如果 $\frac{\eta\gamma}{\delta^2}<\frac{1}{4}$，则存在一个满足 $\|P\|<\frac{2\gamma}{\delta}$ 的 P，使得 $R\left(\begin{pmatrix}I\\P\end{pmatrix}\right)$ 是 A 的不变子空间. 把 $\|P\|<\frac{2\gamma}{\delta}$ 与 (8.6)—(8.9) 联系起来，就刻划了 $R\left(\begin{pmatrix}I\\0\end{pmatrix}\right)$ 作为 A 的近似不变子空间的逼近程度. 这一结论可叙述如下:

定理 8.2 (逼近定理)[152]. 设 $A\in\mathbf{C}^{n\times n}$，$X=(X_1, X_2)$ 是一酉阵，$X_1\in\mathbf{C}^{n\times l}$，$1\leqslant l\leqslant n-1$. 记

$$X^HAX=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},\ A_{11}\in\mathbf{C}^{l\times l}. \tag{8.20}$$

定义算子 $\mathbf{T}$:

$$\mathbf{T}P \equiv PA_{11} - A_{22}P, \quad P \in \mathbf{C}^{(n-l)\times l},$$

并假定 T 可逆. 令 $\delta = \|\mathbf{T}^{-1}\|^{-1}$, $\gamma = \|A_{21}\|$ 和 $\eta = \|A_{12}\|$. 如果

$$\frac{\eta\gamma}{\delta^2} < \frac{1}{4},$$

则存在 $P \in \mathbf{C}^{(n-l)\times l}$, 满足 $\|P\| < \frac{2\gamma}{\delta}$,使 $R\left(X\begin{pmatrix} I \\ P \end{pmatrix}\right)$是$A$的不变子空间.

定理 8.2 表明: $R\left(X\begin{pmatrix} I \\ 0 \end{pmatrix}\right)$(即 $R(X_1)$) 是A的一个近似不变子空间, 并且 $\|P\| < \frac{2\gamma}{\delta}$反映了它和$A$的不变子空间 $R\left(X \cdot \begin{pmatrix} I \\ P \end{pmatrix}\right)$的差距(参看 (8.7)—(8.9)).其中$\gamma$与$\delta$的意义,将在下一段中阐明.

8.3 剩余与矩阵分离度

定理 8.2 中的结论 $\|P\| < \frac{2\gamma}{\delta}$表示: 如果 δ一定, 则 $\gamma = \|A_{21}\|$ 愈小,近似不变子空间 $R(X_1)$ 的近似程度愈高. 我们将利用剩余来说明: $\|A_{21}\|$ 的大小是A的一个近似不变子空间偏离大小的一个合理的测度.

对于 $A \in \mathbf{C}^{n\times n}$, 如果 $R(X_1)$ 是A的不变子空间, 则据定义 8.2, 存在 $A_1 \in \mathbf{C}^{l\times l}$, 使得剩余

$$R \equiv AX_1 - X_1A_1 = 0.$$

如果 $R(X_1)$ 是A的一个近似不变子空间, 则可能取到一个 $A_1 \in \mathbf{C}^{l\times l}$, 使得剩余 $R \equiv AX_1 - X_1A_1$ 的范数很小.

Kahan (1967) 证明了下述结果(参看[155]).

定理 8.3. 设 A, X 与 A_{ij} 皆如定理 8.2 所述. 令

$$R_{A_1} = AX_1 - X_1A_1, \quad A_1 \in \mathbb{C}^{l\times l}.$$

则对于 $\bigcup_{m,n=1}^{N} \mathbb{C}^{m\times n}$ 上任一相容的酉不变范数$\|\quad\|$,有

$$\inf_{A_1\in\mathbb{C}^{l\times l}} \|R_{A_1}\| = \|A_{21}\|.$$

证明: 只需注意到

$$\|R_{A_1}\| = \|X^H R_{A_1}\| = \|X^H A X_1 - X^H X_1 A_1\|$$
$$= \left\|\begin{pmatrix} A_{11} - A_1 \\ A_{21} \end{pmatrix}\right\| \geqslant \|A_{21}\|,$$

并且上式当且仅当 $A_1 = A_{11}$ 时,达到等式. □

下面将说明: 量 $\delta = \|\mathbf{T}^{-1}\|^{-1}$ 在某种意义上反映了 A_{11} 与 A_{22} 的特征值的分离程度.

首先证明

定理 8.4. 设 $A_{11} \in \mathbb{C}^{l\times l}$, $A_{22} \in \mathbb{C}^{(n-l)\times(n-l)}$. 令

$$\mathbf{T}P \equiv PA_{11} - A_{22}P, \quad P \in \mathbb{C}^{(n-l)\times l}.$$

则 $\mathbf{T}$ 的谱

$$\lambda(\mathbf{T}) = \{\lambda - \mu : \lambda \in \lambda(A_{11}), \mu \in \lambda(A_{22})\}; \tag{8.21}$$

因而

$$\mathbf{T}\text{ 可逆} \Longleftrightarrow \lambda(A_{11}) \cap \lambda(A_{22}) = \varnothing, \tag{8.22}$$

并且当 $\mathbf{T}$ 可逆时, 有

$$0 < \delta = \|\mathbf{T}^{-1}\|^{-1} \leqslant \min\{|\lambda - \mu| : \lambda \in \lambda(A_{11}), \mu \in \lambda(A_{22})\}. \tag{8.23}$$

证明:

据定义, $\lambda \in \lambda(\mathbf{T})$ 是指存在 $P \neq 0$, $P \in \mathbb{C}^{(n-l)\times l}$, 使得 $\mathbf{T}P = \lambda P$. 将 P 的诸列连接起来,成为 $\mathbb{C}^{l(n-l)}$ 中的一个向量,于是给出了 $\mathbf{T}$ 的矩阵表示

$$T = A_{11}^T \otimes I^{(n-l)} - I^{(l)} \otimes A_{22}.$$

根据第一章 §5(5.6) 式知,

$$\lambda(T)=\{\lambda-\mu:\lambda\in\lambda(A_{11}),\mu\in\lambda(A_{22})\},$$

因此(8.21)式成立;由此可立即推知(8.22)和(8.23)亦成立. □

(8.23)式说明: $\delta=\|\mathbf{T}^{-1}\|^{-1}$ 反映了 $\lambda(A_{11})$ 与 $\lambda(A_{22})$ 分离的程度. 因此建议引进下述定义.

定义 8.3[152]. 设 $B\in\mathbf{C}^{l\times l}$, $C\in\mathbf{C}^{m\times m}$. 定义算子 $\mathbf{T}:\mathbf{C}^{m\times l}\to\mathbf{C}^{m\times l}$

$$\mathbf{T}P\equiv PB-CP,\quad P\in\mathbf{C}^{m\times l}.$$

则 B 与 C 的分离度 (separation) 定义为

$$\operatorname{sep}(B,C)=\begin{cases}\|\mathbf{T}^{-1}\|^{-1}, & 0\notin\lambda(\mathbf{T})\\ 0, & 0\in\lambda(\mathbf{T}).\end{cases}\tag{8.24}$$

容易验证,下述定义与定义 8.3 等价(本节习题 1).

定义 8.4. 对于定义 8.3 中所述的 B, C 与 $\mathbf{T}$, B 与 C 的分离度定义为

$$\operatorname{sep}(B,C)\equiv\inf_{\|P\|=1}\|\mathbf{T}P\|=\inf_{\|P\|=1}\|PB-CP\|.\tag{8.25}$$

于是,利用函数 sep, 定理 8.2 的结论可以写成

$$\|P\|\leqslant\frac{2\|A_{21}\|}{\operatorname{sep}(A_{11},A_{22})}.\tag{8.26}$$

因此,$\dfrac{1}{\operatorname{sep}(A_{11},A_{22})}$ 可以看作不变子空间的条件数.

以下研究函数 sep 的主要性质.

由(8.23)已知

$$0\leqslant\operatorname{sep}(B,C)\leqslant\min\{|\lambda-\mu|:\lambda\in\lambda(B),\mu\in\lambda(C)\}.$$

更进一步,如果利用 $\|\ \|_F$ 定义 sep, 记为 sep_F, 则可证明

定理 8.5.[20] 设 B 与 C 分别相似变换为 Jordan 标准形的变换矩阵为 Q_B 与 Q_C, 即

$$Q_B^{-1}BQ_B=J_B,\quad Q_C^{-1}CQ_C=J_C;$$

又设 Jordan 标准形 J_B 与 J_C 中最大 Jordan 块的阶数分别为 r_B 与 r_C. 记

$$d(B,C)=\min\{|\lambda-\mu|:\lambda\in\lambda(B),\ \mu\in\lambda(C)\}$$

和

$$\kappa(Q)=\|Q\|_2\|Q^{-1}\|_2.$$

则有

$$\frac{\nu(r_B,r_C)-d(B,C)}{(\nu(r_B,r_C))^{r_B+r_C-1}-(d(B,C))^{r_B+r_C-1}}\cdot\frac{(d(B,C))^{r_B+r_C-1}}{\kappa(Q_B)\kappa(Q_C)}$$
$$\leqslant \mathrm{sep}_F(B,C)\leqslant d(B,C), \tag{8.27}$$

其中

$$\mathrm{sep}_F(B,C)\equiv\inf_{\|P\|_F}\|PB-CP\|_F$$

和

$$\nu(r_B,r_C)=\begin{cases}0 & r_B=r_C=1\\ 1 & r_B=1\text{ 或 }r_C=1,\text{但 }r_Br_C>1\\ 2 & r_B,r_C\geqslant 2.\end{cases}$$

证明见作者的论文[20].

从定理 8.5 可立即得出：当 B 与 C 均为可对角化阵时，有

$$\frac{d(B,C)}{\kappa(Q_B)\kappa(Q_C)}\leqslant \mathrm{sep}_F(B,C)\leqslant d(B,C); \tag{8.28}$$

当 B 与 C 均为正规阵时，有

$$\mathrm{sep}_F(B,C)=d(B,C). \tag{8.29}$$

下述定理说明 $\mathrm{sep}(B,C)$ 对于 B 与 C 的扰动是不敏感的.

定理 8.6.[152] 设 $B,E\in\mathbf{C}^{l\times l},C,F\in\mathbf{C}^{m\times m}$. 则

$$\mathrm{sep}(B,C)-(\|E\|+\|F\|)$$
$$\leqslant \mathrm{sep}(B+E,C+F)\leqslant \mathrm{sep}(B,C)+(\|E\|+\|F\|). \tag{8.30}$$

证明:

关于(8.30)中的第一个不等式,可如下证明.

若 $\mathrm{sep}(B,C)\leqslant\|E\|+\|F\|$, 则不等式显然成立. 因此假定 $\mathrm{sep}(B,C)>\|E\|+\|F\|$. 定义算子 S:

$$\mathrm{S}P=PE-FP,P\in\mathbf{C}^{m\times l}.$$

由 $\|\mathrm{S}\|\leqslant\|E\|+\|F\|$ 可得

$$\|\mathbf{S}\mathbf{T}^{-1}\| \leqslant \frac{\|E\| + \|F\|}{\operatorname{sep}(B,C)} < 1.$$

所以 $I + \mathbf{S}\mathbf{T}^{-1}$ 可逆，从而 $\mathbf{T} + \mathbf{S} = (I + \mathbf{S}\mathbf{T}^{-1})\mathbf{T}$ 可逆．于是

$$\begin{aligned}
\operatorname{sep}(B+E, C+F) &= \|(\mathbf{T}+\mathbf{S})^{-1}\|^{-1} = \|\mathbf{T}^{-1}(I+\mathbf{S}\mathbf{T}^{-1})^{-1}\|^{-1} \\
&\geqslant \|\mathbf{T}^{-1}\|^{-1}\|(I+\mathbf{S}\mathbf{T}^{-1})^{-1}\|^{-1} \geqslant \operatorname{sep}(B,C)(1-\|\mathbf{S}\mathbf{T}^{-1}\|) \\
&\geqslant \operatorname{sep}(B,C)\left(1-\frac{\|E\|+\|F\|}{\operatorname{sep}(B,C)}\right) \\
&= \operatorname{sep}(B,C) - (\|E\|+\|F\|).
\end{aligned}$$

(8.30)的第二个不等式可以从第一个不等式导出．□

定理 8.7[152]．若 B，$X \in \mathbf{C}^{l\times l}$，$C, Y \in \mathbf{C}^{m\times m}$，其中$X$与$Y$皆为非奇异阵．则

$$\operatorname{sep}(XBX^{-1}, YCY^{-1}) \geqslant \frac{\operatorname{sep}(B,C)}{\kappa(X)\kappa(Y)}, \tag{8.31}$$

其中 $\kappa(X) = \|X\|\|X^{-1}\|$，$\kappa(Y) = \|Y\|\|Y^{-1}\|$；若$X$与$Y$均为酉阵，则对于酉不变范数，有

$$\operatorname{sep}(XBX^{-1}, YCY^{-1}) = \operatorname{sep}(B,C). \tag{8.32}$$

证明:

利用(8.25)式，有

$$\begin{aligned}
\operatorname{sep}(XBX^{-1}, YCY^{-1}) &= \inf_{P\neq 0}\frac{\|PXBX^{-1} - YCY^{-1}P\|}{\|P\|} \\
&= \inf_{P\neq 0}\frac{\|Y[(Y^{-1}PX)B - C(Y^{-1}PX)]X^{-1}\|}{\|P\|} \\
&= \inf_{Q\neq 0}\frac{\|Y(QB-CQ)X^{-1}\|}{\|YQX^{-1}\|} \qquad (8.33) \\
&\geqslant \inf_{Q\neq 0}\frac{\|QB-CQ\|}{\|Q\|\|Y\|\|Y^{-1}\|\|X\|\|X^{-1}\|} \\
&= \frac{\operatorname{sep}(B,C)}{\kappa(X)\kappa(Y)}.
\end{aligned}$$

即不等式(8.31)成立．

此外,(8.32)可从(8.33)导出. □

定理 8.8.[152] 若 $B=\operatorname{diag}(B_1,\cdots,B_p)$, $C=\operatorname{diag}(C_1,\cdots,C_q)$, 其中 $B_i(1\leqslant i\leqslant p)$ 与 $C_j(1\leqslant j\leqslant q)$ 均为方阵. 则

$$\operatorname{sep}_F(B,C)=\min\{\operatorname{sep}_F(B_i,C_j):i=1,\cdots,p,\ j=1,\cdots,q\}. \tag{8.34}$$

证明:

只需证明

$$\operatorname{sep}_F\left(B,\begin{pmatrix}C_1&0\\0&C_2\end{pmatrix}\right)=\min\{\operatorname{sep}_F(B,C_1),\operatorname{sep}_F(B,C_2)\}; \tag{8.35}$$

因为同理可证

$$\operatorname{sep}_F\left(\begin{pmatrix}B_1&0\\0&B_2\end{pmatrix},C\right)=\min\{\operatorname{sep}_F(B_1,C),\operatorname{sep}_F(B_2,C)\},$$

然后可以归纳地证明(8.34).

首先指出, $\lambda(B)\cap\lambda\left(\begin{pmatrix}C_1&0\\0&C_2\end{pmatrix}\right)\neq\varnothing$ 的必要与充分条件是 $\lambda(B)\cap\lambda(C_1)\neq\varnothing$ 或 $\lambda(B)\cap\lambda(C_2)\neq\varnothing$. 因此,(8.35)两端必同时为零,或同时不为零. 若同时为零,则等式(8.35)显然成立;所以现在假设(8.35)两端同时不为零. 这时,有

$$\begin{aligned}
\operatorname{sep}_F^2\left(B,\begin{pmatrix}C_1&0\\0&C_2\end{pmatrix}\right)&=\inf_{\|P\|_F=1}\left\|\begin{pmatrix}P_1\\P_2\end{pmatrix}B-\begin{pmatrix}C_1&0\\0&C_2\end{pmatrix}\begin{pmatrix}P_1\\P_2\end{pmatrix}\right\|_F^2\\
&=\inf_{\|P\|_F=1}\left\|\begin{pmatrix}P_1B-C_1P_1\\P_2B-C_2P_2\end{pmatrix}\right\|_F^2\\
&=\inf_{\|P\|_F=1}(\|P_1B-C_1P_1\|_F^2+\|P_2B-C_2P_2\|_F^2)\\
&\geqslant\inf_{\|P_1\|_F^2+\|P_2\|_F^2=1}(\operatorname{sep}_F^2(B,C_1)\|P_1\|_F^2+\operatorname{sep}_F^2(B,C_2)\|P_2\|_F^2)\\
&\geqslant\min\{\operatorname{sep}_F^2(B,C_1),\operatorname{sep}_F^2(B,C_2)\}.
\end{aligned}$$

另一方面, 无妨设 $\operatorname{sep}_F(B,C_1)\leqslant\operatorname{sep}_F(B,C_2)$. 取 P_1 满足 $\|P_1\|_F=1$, 并且

$$\text{sep}_F(B, C_1) = \|P_1B - C_1P_1\|_F.$$

则

$$\text{sep}_F(B, C) = \inf_{\|P\|_F=1} \|PB - CP\|_F \leqslant \left\| \begin{pmatrix} P_1 \\ 0 \end{pmatrix} B - \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix} \begin{pmatrix} P_1 \\ 0 \end{pmatrix} \right\|_F$$

$$= \|P_1B - C_1P_1\|_F = \text{sep}_F(B, C_1). \quad \square$$

定理 8.9[152] 设范数 $\|\ \ \|_r$ 与 $\|\ \ \|_s$ 满足

$$\sigma\|P\|_r \leqslant \|P\|_s \leqslant \tau\|P\|_r, \forall P \in \mathbf{C}^{m\times l}.$$

则

$$\text{sep}_r(B, C) \geqslant \frac{\sigma}{\tau}\text{sep}_s(B, C), \forall B \in \mathbf{C}^{l\times l}, \forall C \in \mathbf{C}^{m\times m}. \quad (8.36)$$

其中

$$\text{sep}_r(B, C) \equiv \inf_{\|P\|_r=1} \|PB - CP\|_r,$$

$$\text{sep}_s(B, C) \equiv \inf_{\|P\|_s=1} \|PB - CP\|_s.$$

证明:

首先有不等式

$$\tau\|PB - CP\|_r \geqslant \|PB - CP\|_s \geqslant \text{sep}_s(B, C)\|P\|_s$$
$$\geqslant \sigma\text{sep}_s(B, C)\|P\|_r.$$

因此,对于 $P \neq 0$, 有

$$\frac{\|PB - CP\|_r}{\|P\|_r} \geqslant \frac{\sigma}{\tau}\text{sep}_s(B, C). \quad (8.37)$$

在(8.37)式左端取 $\|P\|_r = 1$ 上的极小值,即给出不等式(8.36). $\square$

在不变子空间扰动理论中,$\dfrac{1}{\text{sep}(A_{11}, A_{22})}$ 作为不变子空间的条件数(参看(8.26)式),是一个相当重要的量; 函数 sep 的上述性质(如定理 8.5 一定理 8.9 所述),说明它包含着丰富的内容. 可是如何估算 $\text{sep}(A_{11}, A_{22})$, 却是一个困难的问题. 关于这方面的研究,可参看 Varah [178] 和本书作者的论文[20]与[21].

8.4 扰动定理

一个逼近定理，比如定理 8.2，可以转化成一个扰动定理，这只需要把 A 的不变子空间看作 $A+E$ 的近似不变子空间．因此，由定理 8.2 和定理 8.6，立即得到下述结果．

定理 8.10 （扰动定理）[152]．设 $A,E\in\mathbb{C}^{n\times n}$．$X=(X_1,X_2)$ 为酉阵，$X_1\in\mathbb{C}^{n\times l}(1\leqslant l\leqslant n-1)$，并假定 $R(X_1)$ 是 A 的不变子空间．又设 X^HAX 和 X^HEX 与 X 相一致地分块为

$$X^HAX=\begin{pmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{pmatrix},\ X^HEX=\begin{pmatrix}E_{11} & E_{12}\\ E_{21} & E_{22}\end{pmatrix}.$$

如果

$$\hat{\delta}\equiv\operatorname{sep}(A_{11},A_{22})-(\|E_{11}\|+\|E_{22}\|)>0 \tag{8.38}$$

并且

$$\frac{\|E_{21}\|(\|A_{12}\|+\|E_{12}\|)}{\hat{\delta}^2}<\frac{1}{4}, \tag{8.39}$$

则存在矩阵 $P\in\mathbb{C}^{(n-l)\times l}$，满足

$$\|P\|\leqslant\frac{2\|E_{21}\|}{\hat{\delta}}, \tag{8.40}$$

使得 $R\left(X\begin{pmatrix}I\\ P\end{pmatrix}\right)$ 是 $A+E$ 的不变子空间．

下面用一个最简单的情况，来说明定理 8.10 的结论．

推论 8.1．设 $A,E\in\mathbb{C}^{n\times n}$．$\lambda_1$ 是 A 的单特征值，$\boldsymbol{x}_1$ 是相应的特征向量，并且 $\|\boldsymbol{x}_1\|_2=1$．设 $X=(\boldsymbol{x}_1,X_2)$ 是一酉阵，使得

$$X^HAX=\begin{pmatrix}\lambda_1 & \boldsymbol{a}^H\\ 0 & A_{22}\end{pmatrix},\quad X^HEX=\begin{pmatrix}e_{11} & \boldsymbol{f}^H\\ \boldsymbol{g} & E_{22}\end{pmatrix}.$$

令

$$\delta=\|(\lambda_1 I-A_{22})^{-1}\|_2^{-1},\ \varepsilon=\|E\|_2,\ \eta=\|\boldsymbol{a}\|_2,$$
$$\gamma=\|\boldsymbol{g}\|_2.$$

如果

$$\delta-2\varepsilon>0 \tag{8.41}$$

并且

$$\frac{\gamma(\eta+\varepsilon)}{(\delta-2\varepsilon)^2}<\frac{1}{4}, \tag{8.42}$$

则存在 $\boldsymbol{p}\in\mathbf{C}^{n-1}$，满足 $\|\boldsymbol{p}\|_2\leqslant\frac{2\gamma}{\delta-2\varepsilon}$，使得 $\tilde{\boldsymbol{x}}_1=\boldsymbol{x}_1+X_2\boldsymbol{p}$ 张成$A+E$的不变子空间. $R(\tilde{\boldsymbol{x}}_1)$ 与 $R(\boldsymbol{x}_1)$ 之间的距离

$$\sin\theta(R(\tilde{\boldsymbol{x}}_1),R(\boldsymbol{x}_1))\leqslant\frac{2\gamma}{\sqrt{(\delta-2\varepsilon)^2+4\gamma^2}}. \tag{8.43}$$

证明:

直接应用定理 8.10. 只需注意两点:

1) $\mathrm{sep}_2(\lambda_1,A_{22})=\|(\lambda_1I-A_{22})^{-1}\|_2^{-1}$, $\quad$ (8.44)

这是因为,对于如下定义的算子 $\mathbf{T}$:

$$\mathbf{T}\boldsymbol{p}\equiv\boldsymbol{p}\lambda_1-A_{22}\boldsymbol{p}=(\lambda_1I-A_{22})\boldsymbol{p},\ \boldsymbol{p}\in\mathbf{C}^{n-1},$$

显然有

$$\|\mathbf{T}^{-1}\|_2^{-1}=\|(\lambda_1I-A_{22})^{-1}\|_2^{-2}.$$

2) 由(8.8)式知

$$\sin\theta(R(\tilde{\boldsymbol{x}}_1),R(\boldsymbol{x}_1))=\frac{\|\boldsymbol{p}\|_2}{\sqrt{1+\|\boldsymbol{p}\|_2^2}},$$

再将 $\|\boldsymbol{p}\|_2\leqslant\frac{2\gamma}{\delta-2\varepsilon}$ 代入上式,便得到(8.43). □

注 8.1. 因为推论 8.1 中的 $\gamma=\|\boldsymbol{g}\|_2\leqslant\|E\|_2=\varepsilon$，所以由(8.43)可以得到

$$\sin\theta(R(\tilde{\boldsymbol{x}}_1),R(\boldsymbol{x}_1))\leqslant\frac{2\varepsilon}{\sqrt{(\delta-2\varepsilon)^2+4\varepsilon^2}}. \tag{8.45}$$

(8.45)式表明: 在扰动量(即指 ε)相同的条件下，一般说来，$\delta=\mathrm{sep}_2(\lambda_1,A_{22})$ 愈大,则 A 的不变子空间 $R(\boldsymbol{x}_1)$ 愈稳定.

但是必须指出一个重要事实: **即使 λ_1 与 $\lambda(A_{22})$ 明显分离，$\mathrm{sep}_2(\lambda_1,A_{22})$ 也可能非常之小.**

例 8.2.

$$\lambda_1 = 1 + \alpha, A_{22} = \begin{pmatrix} 1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 1 \end{pmatrix} \in \mathbf{C}^{(n-1)\times(n-1)},\ 0 < \alpha < 1.$$

由 (8.44) 可知，如果令 $J_{n-1}(\alpha) = \begin{pmatrix} \alpha & -1 & & \\ & \ddots & \ddots & \\ & & \ddots & -1 \\ & & & \alpha \end{pmatrix} \in \mathbf{C}^{(n-1)\times(n-1)}$，则

$$\begin{aligned} \mathrm{sep}_2(\lambda_1, A_{22}) &= \|J_{n-1}^{-1}(\alpha)\|_2^{-1} = \sigma_{\min}(J_{n-1}(\alpha)) \\ &= \inf_{\substack{v \neq 0 \\ v \in \mathbf{C}^{n-1}}} \frac{\|J_{n-1}(\alpha)\boldsymbol{v}\|_2}{\|\boldsymbol{v}\|_2}. \end{aligned}$$

上式中的 $\sigma_{\min}(J_{n-1}(\alpha))$ 表示 $J_{n-1}(\alpha)$ 的最小奇异值. 今取 $\boldsymbol{v} = (1, \alpha, \cdots, \alpha^{n-2})^T$，有

$$\begin{aligned} \frac{\|J_{n-1}(\alpha)\boldsymbol{v}\|_2}{\|\boldsymbol{v}\|_2} &= \frac{\|(0, \cdots, 0, \alpha^{n-1})^T\|_2}{\|(1, \alpha, \cdots, \alpha^{n-2})^T\|_2} \\ &= \frac{\alpha^{n-1}}{\sqrt{1 + \alpha^2 + \cdots + \alpha^{2(n-2)}}} < \alpha^{n-1}, \end{aligned}$$

因此

$$\mathrm{sep}_2(\lambda_1, A_{22}) < \alpha^{n-1}.$$

这个例子说明，对于比较大的 n 和一个适度大小的 $\alpha \in (0,1)$，尽管 λ_1 与 A_{22} 的特征值之间的距离为 α，但 λ_1 与 A_{22} 的分离度却非常之小(小于 α^{n-1}).

习题

1. 设 $B \in \mathbf{C}^{l\times l}$，$C \in \mathbf{C}^{m\times m}$. 试证 B 与 C 的分离度的定义 8.3 与定义 8.4 等价.

2. 设 $A, \widetilde{A} \in \mathbf{C}^{n\times n}$，$X = (X_1, X_2)$ 与 $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2)$ 为酉阵，其中 $X_1, \widetilde{X}_1 \in \mathbf{C}^{n\times l}$. 又设

$$X^{-1}AX = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \widetilde{X}^{-1}\widetilde{A}\widetilde{X} = \begin{pmatrix} \widetilde{A}_{11} & \widetilde{A}_{12} \\ 0 & \widetilde{A}_{22} \end{pmatrix},$$

其中 $A_{11}, \widetilde{A}_{11} \in \mathbf{C}^{l\times l}(1 \leqslant l \leqslant n-1)$. 令 $\mathscr{X} = R(X_1)$, $\widetilde{\mathscr{X}} = R(\widetilde{X}_1)$ 和 $R = \widetilde{A}X_1 - X_1A_{11}$. 试证: 如果 $\lambda(A_{11}) \cap \lambda(\widetilde{A}_{22}) = \emptyset$, 则对于 $\mathbf{C}^{l\times l}$ 上的任一酉不变范数$\|\ \|$,必有

$$\|\sin\Theta(\mathscr{X}, \widetilde{\mathscr{X}})\| \leqslant \frac{\|R\|}{\operatorname{sep}(A_{11}, \widetilde{A}_{22})}.$$

第三章说明

本章讨论矩阵特征值的扰动和矩阵的不变子空间（包括特征向量所张成的子空间)的扰动. 这方面的文献十分丰富,但限于篇幅,作者只能挑选出一些重要结果加以阐述.

关于 Hermite 阵特征值扰动性质的研究，已经相当深入，读者可参看文献[39]、[40]、[115]、[132]、[148]、[187] 和 [197].关于正规矩阵特征值的扰动，最近， Bhatia、Davis 和 McIntosh 的论文 [58] 证明了一些新的结果；此外，本书作者的论文 [19] 推广了 Hoffman-Wielandt 定理.

对于一般矩阵， Ostrowski [130] 最早证明了矩阵特征值的连续性，并且给出了可以计算的扰动上界. 之后，Henrici 和 Bhatia 等人,也得到了若干可以计算的上界(参看[59]、[60]、[86] 和[102]). 有一些关于矩阵特征值扰动的定理（例如 Bauer-Fike 定理及其推广),在扰动界限中包含着不易计算的量. 如何估算出这些不易算出的量,是值得进一步研究的课题(关于谱条件数的估算,可参看文献[20]、[21]、[33]和[150]).

关于不变子空间的扰动，本章介绍了 Davis-Kahan 理论和 Stewart 理论. Davis 与 Kahan [75] 和 Stewart [152] 从不同角度研究了不变子空间的扰动性质. 值得指出的是，他们所研究的仅仅是一些特殊类型的不变子空间；对于更广一类的不变子空间 (Bart 和 Gohberg 等人称之为稳定的不变子空间,见[42])的扰动性质,至今还缺乏研究.

关于奇异值的扰动,除了本章 §3 定理 3.10—3.12 这些基本结

果之外，读者还可参看文献[5]、[160]和[162]. 关于与奇异值分解相联系的一些子空间的扰动，可参看[155]和[182].

对于非负矩阵的特征值及其扰动性质，本书未作介绍；请读者参看文献[62]、[68]、[92]和[146].

第四章　广义特征值问题扰动分析

§1　基本概念

这里所说的广义特征值问题，是指线性广义特征值问题

$$Ax = \lambda Bx, \tag{1.1}$$

其中 $A, B \in \mathbb{C}^{m\times n}$, $x \in \mathbb{C}^n$, $x \neq 0$, $\lambda \in \mathbb{C}$.

这个问题的重要性，最初是从常微分方程组

$$B\dot{x} = Ax,\ x = (\xi_1(t), \cdots, \xi_n(t))^T \tag{1.2}$$

的求解看出来的，此处 $A, B \in \mathbb{R}^{n\times n}$, $\xi_i(t)\ (1 \leqslant i \leqslant n)$ 是 $t \in \mathbb{R}$ 的实值函数. 如果 $\lambda \in \mathbb{C}$ 与非零向量 $u \in \mathbb{C}^n$ 满足

$$Au = \lambda Bu$$

(即 λ 与 u 分别是 $Au = \lambda Bu$ 的特征值与特征向量)，则

$$x = ue^{\lambda t}$$

是方程组(1.2)的一个解.

随着科学研究的深入，数学物理方程、统计分析和自动控制等方面的许多问题的数值求解，往往归结为(1.1)所示的广义特征值问题的研究.

广义特征值问题(1.1)的代数理论（即矩阵对的标准形理论），在十九世纪已经为 Weierstrass 和 Kronecker 所建立. 可是从数值分析的角度，把广义特征值问题作为一个独立的课题进行研究，至今只有十几年的历史. 主要的工作是从七十年代开始的.

在 Wilkinson 和 Peters (1970) 发表关于广义特征值问题解法的论文之后，出现了一系列的论文，讨论 $Ax = \lambda Bx$ 的特征值与特征向量的计算方法，比较著名的有 Fix-Heiberger 算法 (1972), Moler-Stewart QZ 算法 (1973) 和 Crawford 算法 (1973, 1976), 等等.

几乎与研究广义特征值问题(1.1)算法的同时，开始了扰动理论的研究. Stewart (1972) 的论文,可以作为这一研究的起点. 随后， Stewart (1973, 1975, 1979), Fix, Heiberger (1972) 和 Crawford (1976) 发表了他们在这方面的结果. 七十年代末以来，作者在这方面进行了比较系统的研究；收入《Lecture Notes in Math. 973》(1983) 的论文 "*Perturbation analysis for the generalized eigenvalue and the generalized singular value problem*"，是作者前一阶段工作的小结.

值得指出的是，广义特征值问题 (1.1)，无论从算法方面还是从扰动分析方面,都不能简单地归结为普通特征值问题 $A\boldsymbol{x}=\lambda\boldsymbol{x}$. 因为，即使在 A 与 B 都是方阵的情形，有可能 A 与 B 均为奇异阵，或者是近乎奇异的矩阵,对它们无法进行求逆的运算,或者它们的求逆是不稳定的. 另外，在下面我们还将会看到，广义特征值问题(1.1)的特征值不是分布在复平面 $\mathbf{C}$ 内,而是分布在复投影平面 $\mathbf{G}_{1,2}$ 上. 因此,有必要把广义特征值问题(1,1)作为一个独立的课题进行讨论.

1.1 正则对与奇异对

对于 $A,B\in\mathbf{C}^{m\times n}$，通常把 $A+\lambda B$ 叫做矩阵束,其中 $\lambda\in\mathbf{C}$. 与矩阵束 $A+\lambda B$ 相对应的一对矩阵 A 与 B,叫做矩阵对，记作 $\{A, B\}$.

定义 1.1. 设 $A,B\in\mathbf{C}^{m\times n}$. 如果 $m=n$，并且

$$\det(A+\lambda B)\not\equiv 0,\quad \lambda\in\mathbf{C},$$

则称 $\{A,B\}$ 为 n 阶正则矩阵对(以下简称为 n 阶正则对)；如果 $m\neq n$，或者 $m=n$ 但

$$\det(A+\lambda B)\equiv 0,\quad \lambda\in\mathbf{C},$$

则称 $\{A,B\}$ 为奇异矩阵对(以下简称为奇异对).

Weierstrass 和 Kronecker 分别给出了正则对与奇异对的标准形. 下面的定理 1.1 和定理 1.2 叙述了他们的结果,定理的证明可在 Гантмахер 的《矩阵论》一书 (即文献[197])中找到.

定理 1.1 (Weierstrass 1867). 设 $\{A,B\}$ 是 n 阶正则对. 则必存在非奇异阵 $P,Q\in\mathbb{C}^{n\times n}$, 使得

$$PAQ=J_A,\quad PBQ=J_B. \tag{1.3}$$

其中

$$J_A=\begin{pmatrix}J & 0\\ 0 & I^{(n_2)}\end{pmatrix},\quad J_B=\begin{pmatrix}I^{(n_1)} & 0\\ 0 & N\end{pmatrix}; \tag{1.4}$$

$$\left.\begin{aligned}&J=\operatorname{diag}(J_1(\lambda_1),\cdots,J_r(\lambda_r))\in\mathbb{C}^{n_1\times n_1},\\ &\lambda_i\neq\lambda_j(i\neq j),1\leqslant i,j\leqslant r,\end{aligned}\right\} \tag{1.5}$$

$$\left.\begin{aligned}&J_i(\lambda_i)=\operatorname{diag}(J_i^{(1)}(\lambda_i),\cdots,J_i^{(k_i)}(\lambda_i))\in\mathbb{C}^{n(\lambda_i)\times n(\lambda_i)},\\ &J_i^{(k)}(\lambda_i)=\begin{pmatrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}\in\mathbb{C}^{n_k(\lambda_i)\times n_k(\lambda_i)},\\ &\qquad\qquad 1\leqslant k\leqslant k_i,\ 1\leqslant i\leqslant r,\end{aligned}\right\} \tag{1.6}$$

$$\left.\begin{aligned}&\sum_{k=1}^{k_i}n_k(\lambda_i)=n(\lambda_i),\quad 1\leqslant i\leqslant r,\\ &\sum_{i=1}^{r}n(\lambda_i)=n_1;\end{aligned}\right\} \tag{1.7}$$

$$N=\operatorname{diag}(N^{(l_1)},\cdots,N^{(l_s)})\in\mathbb{C}^{n_2\times n_2}, \tag{1.8}$$

$$\left.\begin{aligned}&N^{(l_j)}=\begin{pmatrix}0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & 0\end{pmatrix}\in\mathbb{C}^{l_j\times l_j},\quad 1\leqslant j\leqslant s,\\ &\sum_{i=1}^{s}l_i=n_2\end{aligned}\right\} \tag{1.9}$$

和

$$n_1+n_2=n.$$

定理 1.2 (Kronecker, 1890). 设 $A,B\in\mathbb{C}^{m\times n}$, $\{A,B\}$ 为奇异对. 则必存在非奇异阵 $P\in\mathbb{C}^{m\times m}$ 和 $Q\in\mathbb{C}^{n\times n}$, 使得

$$PAQ=\begin{pmatrix}J_A & 0\\ 0 & L_A\end{pmatrix},\quad PBQ=\begin{pmatrix}J_B & 0\\ 0 & L_B\end{pmatrix}. \tag{1.10}$$

其中 $J_A, J_B \in \mathbf{C}^{(n_1+n_2)\times(n_1+n_2)}$，如 (1.4)—(1.9) 式所示，即 $\{J_A, J_B\}$ 为正则对；此外，L_A 与 L_B 的形式如下：

$$\left.\begin{aligned}
&L_A = \begin{pmatrix} L_{f_1} & & & & \\ & \ddots & & & \\ & & L_{f_p} & & \\ & & & L_{g_1}^T & \\ & & & & \ddots \\ & & & & & L_{g_q}^T \end{pmatrix}, \quad L_B = \begin{pmatrix} L'_{f_1} & & & & \\ & \ddots & & & \\ & & L'_{f_p} & & \\ & & & L'^T_{g_1} & \\ & & & & \ddots \\ & & & & & L'^T_{g_q} \end{pmatrix}, \\
&L_k = \begin{pmatrix} 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix} \in \mathbf{C}^{k\times(k+1)}, \\
&L'_k = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \end{pmatrix} \in \mathbf{C}^{k\times(k+1)} \\
&k = f_1, \cdots, f_p, g_1, \cdots, g_q.
\end{aligned}\right\} \tag{1.11}$$

1.2 特征值与特征向量

设 $\{A, B\}$ 是 n 阶正则对. 通常把适合 (1.1) 的 $\lambda \in \mathbf{C}$ 与非零向量 $\boldsymbol{x} \in \mathbf{C}^n$ 定义为广义特征值问题 (1.1) 的特征值与特征向量. 按照这样的定义，问题(1.1)的特征值应是方程

$$\det(A - \lambda B) = 0 \tag{1.12}$$

的解 λ. 值得注意的是，当 B 为奇异阵时，(1.12) 式左端多项式的次数低于 n，因而方程式 (1.12) 解的个数少于 n. 不过，如果将 $\lambda = \frac{1}{\mu}$ 代入(1.1)式，则由

$$B\boldsymbol{x} = \mu A\boldsymbol{x} \tag{1.13}$$

可知，$\mu = 0$ 是广义特征值问题 (1.13) 的特征值，因此 $\lambda = \infty$ 也是(1.1)的特征值. 这就是说，广义特征值问题(1.1)的特征值不是分布在 $\mathbf{C}$ 内，而是分布在 $\mathbf{C} \cup \{\infty\}$ 上.

例如，$A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$，$B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. $\{A, B\}$ 显然是正则

对，并且容易看出 $Ax=\lambda Bx$ 有特征值 $\lambda_1=0$ 和 $\lambda_2=\infty$，相应的特征向量为 $x_1=(1,0)^T$ 和 $x_2=(0,1)^T$.

根据投影几何学的理论，$\mathbf{C}\cup\{\infty\}$ 上的每个点 λ，可以用齐次坐标$(\alpha,\beta)\neq(0,0)$表示，当 $\beta\neq 0$ 时，(α,β) 表示一个有穷点 $\lambda=\alpha/\beta$，当 $\beta=0$ 时，(α,β) 表示无穷远点∞. 于是，$\mathbf{C}\cup\{\infty\}$ 就成了复投影平面 $\mathbf{G}_{1,2}$，即 $\mathbf{C}^2$ 内所有 1 维行空间（即 1 维行向量所张成的空间）的全体（见第二章 § 4）.

作为 $\mathbf{G}_{1,2}$ 的一种表示方法，可以记

$$\mathbf{G}_{1,2}=\{(\alpha,\beta)\neq(0,0):\alpha,\beta\in\mathbf{C}\}. \tag{1.14}$$

但必须说明，(1.14) 式中的 (α,β) 是一个等价类中的代表元素，即：如果存在非零复数 ω，使得 $(\alpha_1,\beta_1)=\omega(\alpha,\beta)$，则 (α_1,β_1) 与 (α,β) 表示 $\mathbf{G}_{1,2}$ 上的同一个点. 当 $\alpha\neq 0$ 时，$(\alpha,0)$ 表示无穷远点；当 $\beta\neq 0$ 时，点 (α,β) 的非齐次坐标为 $\lambda=\alpha/\beta\in\mathbf{C}$.

利用 λ 的齐次坐标 (α,β)，广义特征值问题 (1.1) 可以写成比较合理、而且更便于研究的形式

$$\beta Ax=\alpha Bx. \tag{1.15}$$

本章就是从 (1.15) 式出发，讨论广义特征值问题的扰动理论.

定义 1.2. 设$\{A,B\}$为 n 阶正则对. 如果存在 $(\alpha,\beta)\in\mathbf{G}_{1,2}$ 和非零向量 $x\in\mathbf{C}^n$，使得(1.15)式成立，则称 (α,β) 为 $\{A,B\}$ 的特征值，x 叫做 $\{A,B\}$ 属于 (α,β) 的特征向量.

$\{A,B\}$的所有特征值的全体，叫做$\{A,B\}$的谱，记作 $\lambda(A,B)$.

由定义 1.2 可知，

$$\lambda(A,B)=\{(\alpha,\beta)\in\mathbf{G}_{1,2}:\det(\beta A-\alpha B)=0\}. \tag{1.16}$$

因此，有

推论 1.1. 设 $\{A,B\}$ 为 n 阶正则对. 如果 $P,Q\in\mathbf{C}^{n\times n}$ 为非奇异阵，则

$$\lambda(PAQ,PBQ)=\lambda(A,B).$$

此外，可证

推论 1.2. 设 $\{A,B\}$ 为 n 阶正则对. 则 $\{A,B\}$ 恰有 n 个特征值.

证明:

因为 $\{A,B\}$ 为正则对,所以存在 $(\tau,\sigma)\in\mathbf{G}_{1,2}$，并且无妨假设 $|\tau|^2+|\sigma|^2=1$，使得

$$\det(\tau A+\sigma B)\neq 0.$$

作变换

$$A'=-\bar{\sigma}A+\bar{\tau}B,\quad B'=\tau A+\sigma B \tag{1.17}$$

即

$$A=-\sigma A'+\bar{\tau}B',\quad B=\tau A'+\bar{\sigma}B'.$$

则特征值问题(1.15)可改写成等价的形式

$$\beta'A'\boldsymbol{x}=\alpha'B'\boldsymbol{x}, \tag{1.18}$$

其中

$$\alpha'=-\bar{\sigma}\alpha+\bar{\tau}\beta,\ \beta'=\tau\alpha+\sigma\beta. \tag{1.19}$$

注意到 B' 为非奇异阵,因而 $\lambda(A',B')\subset\mathbf{C}$，并且特征值问题(1.18)就是普通特征值问题

$$\tilde{A}\boldsymbol{x}=\tilde{\lambda}\boldsymbol{x},$$

其中

$$\tilde{A}=B'^{-1}A',\ \tilde{\lambda}=\frac{\alpha'}{\beta'}.$$

显然 $\tilde{A}$ 恰有 n 个特征值,假设它们是 $\tilde{\lambda}_1,\cdots,\tilde{\lambda}_n$，则 $\lambda(A',B')=\{(\tilde{\lambda}_i,1)\}_{i=1}^n$. 于是利用变换(1.19)可知 $\{A,B\}$ 亦恰有 n 个特征值 $(\alpha_i,\beta_i),i=1,\cdots,n$，其中

$$\alpha_i=-\sigma\tilde{\lambda}_i+\bar{\tau},\quad \beta_i=\tau\tilde{\lambda}_i+\bar{\sigma}.$$ □

把 Schur 定理(第一章定理 1.2)推广到矩阵对，有下述结论成立.

定理 1.3 (Stewart, [153]). 设 $\{A,B\}$ 为 n 阶正则对. 则存在酉阵 $U,V\in\mathbf{C}^{n\times n}$，使得

$$U^HAV=\begin{pmatrix}\alpha_1 & & *\\ & \ddots & \\ 0 & & \alpha_n\end{pmatrix}\equiv T_A,\ U^HBV=\begin{pmatrix}\beta_1 & & *\\ & \ddots & \\ 0 & & \beta_n\end{pmatrix}\equiv T_B. \tag{1.20}$$

其中 (α_i,β_i) $(i=1,\cdots,n)$ 是 $\{A,B\}$ 的特征值，并且可以适当选取 U 与 V，使得 T_A 与 T_B 的对角线元素，按任一指定的顺

序排列.

证明:

对 A 与 B 的阶数 n 应用数学归纳法. 当 $n=1$ 时,定理1.3的结论显然成立. 假定定理的结论对于 $n-1$ 阶正则对已经成立,现在考虑 n 阶正则对 $\{A,B\}$.

设 (α,β) 是事先规定的特征值顺序中的第一个特征值,$\boldsymbol{v}_1$ 是 $\{A,B\}$ 属于 (α,β) 的单位特征向量,即

$$\beta A\boldsymbol{v}_1=\alpha B\boldsymbol{v}_1. \tag{1.21}$$

容易看出$(A\boldsymbol{v}_1,B\boldsymbol{v}_1)\neq(0,0)$ (如果 $A\boldsymbol{v}_1=B\boldsymbol{v}_1=0$, 则因 $\{A,B\}$ 是正则对, 存在 $\lambda\in\mathbf{C}$, 使得 $\det(A+\lambda B)\neq 0$, 从而由 $(A+\lambda B)\boldsymbol{v}_1=0$ 导出 $\boldsymbol{v}_1=0$; 这与 $\|\boldsymbol{v}_1\|_2=1$ 矛盾),并且(1.21)式表明,存在 $\mathbf{C}^n$ 的1维子空间 $\mathscr{X}_1$, 使得

$$AR(\boldsymbol{v}_1)+BR(\boldsymbol{v}_1)=\mathscr{X}_1,$$

此处 $R(\boldsymbol{v}_1)$ 表示由 $\boldsymbol{v}_1$ 张成的子空间. 取 $V_2\in\mathbf{C}^{n\times(n-1)}$, 使 $V_0=(\boldsymbol{v}_1,V_2)$ 为酉阵; 同时在 $\mathscr{X}_1$ 内取单位向量 $\boldsymbol{u}_1$, 再取 $U_2\in\mathbf{C}^{n\times(n-1)}$, 使 $U_0=(\boldsymbol{u}_1,U_2)$ 为酉阵. 于是,有

$$U_2^HA\boldsymbol{v}_1=0,\quad U_2^HB\boldsymbol{v}_1=0,$$

即

$$U_0^HAV_0=\begin{pmatrix}\boldsymbol{u}_1^HA\boldsymbol{v}_1 & * \\ 0 & A_1\end{pmatrix},\ U_0^HBV_0=\begin{pmatrix}\boldsymbol{u}_1^HB\boldsymbol{v}_1 & * \\ 0 & B_1\end{pmatrix}.$$

因为 $\{A,B\}$ 是正则对,所以 $(\boldsymbol{u}_1^HA\boldsymbol{v}_1,\boldsymbol{u}_1^HB\boldsymbol{v}_1)\neq(0,0)$. 由

$$\beta\boldsymbol{u}_1^HA\boldsymbol{v}_1=\alpha\boldsymbol{u}_1^HB\boldsymbol{v}_1$$

可知, 如果记

$$\alpha_1=\boldsymbol{u}_1^HA\boldsymbol{v}_1,\quad \beta_1=\boldsymbol{u}_1^HB\boldsymbol{v}_1,$$

则存在 $\tau\in\mathbf{C}$, $\tau\neq 0$, 满足

$$(\alpha_1,\beta_1)=\tau(\alpha,\beta),$$

即 (α_1,β_1) 与 (α,β) 表示 $\mathbf{G}_{1,2}$ 上的同一个点. 所以 (α_1,β_1) 就是事先规定顺序中的第一个特征值. 因此,有

$$U_0^HAV_0=\begin{pmatrix}\alpha_1 & * \\ 0 & A_1\end{pmatrix},\quad U_0^HBV_0=\begin{pmatrix}\beta_1 & * \\ 0 & B_1\end{pmatrix}.$$

然后应用归纳法假设，对于 $n-1$ 阶正则对 $\{A_1, B_1\}$，存在 $n-1$ 阶酉阵 U_1 与 V_1，使得

$$U_1^H A_1 V_1 = \begin{pmatrix} \alpha_2 & & * \\ & \ddots & \\ 0 & & \alpha_n \end{pmatrix}, \quad U_1^H B_1 V_1 = \begin{pmatrix} \beta_2 & & * \\ & \ddots & \\ 0 & & \beta_n \end{pmatrix},$$

其中 $(\alpha_2, \beta_2), \cdots, (\alpha_n, \beta_n)$ 按事先规定的顺序排列.

最后，令 $U = \begin{pmatrix} 1 & 0 \\ 0 & U_1 \end{pmatrix} U_0$ 与 $V = \begin{pmatrix} 1 & 0 \\ 0 & V_1 \end{pmatrix} V_0$. 显然 U 与 V 均为酉阵，并且使 (1.20) 式成立. □

从定理 1.3 立即得出

推论 1.3. 设 $\{A, B\}$ 为 n 阶正则对. 则存在非奇异阵 P 与酉阵 V，使得

$$A = P\begin{pmatrix} \alpha_1 & & * \\ & \ddots & \\ 0 & & \alpha_n \end{pmatrix} V^H, \quad B = P\begin{pmatrix} \beta_1 & & * \\ & \ddots & \\ 0 & & \beta_n \end{pmatrix} V^H, \qquad (1.22)$$

其中 (α_i, β_i) 满足 $|\alpha_i|^2 + |\beta_i|^2 = 1$, $i = 1, \cdots, n$.

定义 1.3[14]. 设 $\{A, B\}$ 为奇异对，其中 $A, B \in \mathbf{C}^{m\times n}$. 如果

$$\max_{(\lambda,\mu)\in \mathbf{G}_{1,2}} \operatorname{rank}(\mu A - \lambda B) = k,$$

而 $(\alpha, \beta) \in \mathbf{G}_{1,2}$ 使得

$$\operatorname{rank}(\beta A - \alpha B) < k,$$

则称 (α, β) 是 $\{A, B\}$ 的一个特征值. 如果非零向量 $\boldsymbol{x} \in \mathbf{C}^n$，使得

$$\beta A\boldsymbol{x} = \alpha B\boldsymbol{x}, \quad (A\boldsymbol{x}, B\boldsymbol{x}) \neq (0, 0),$$

则称 $\boldsymbol{x}$ 是 $\{A, B\}$ 属于特征值 (α, β) 的特征向量.

$\{A, B\}$ 的所有特征值的全体，记作 $\lambda(A, B)$.

容易看出，定义 1.3 是定义 1.2 的推广，并且有类似于推论 1.1 的结论，即有

推论 1.4. 设 $\{A, B\}$ 为奇异对，其中 $A, B \in \mathbf{C}^{m\times n}$. 如果 $P \in \mathbf{C}^{m\times m}$ 与 $Q \in \mathbf{C}^{n\times n}$ 均为非奇异阵，则

$$\lambda(PAQ, PBQ) = \lambda(A, B).$$

此外，还有

推论 1.5. 设 $\{A,B\}$ 为奇异对，A 与 B 有(1.10)式所示的标准形. 则

$$\lambda(A,B)=\lambda(J_A,J_B).$$

证明:

利用推论 1.3 和标准形 (1.10) 可知

$$\begin{aligned}\operatorname{rank}(\mu A-\lambda B)&=\operatorname{rank}\left(\mu\begin{pmatrix}J_A & 0\\ 0 & L_A\end{pmatrix}-\lambda\begin{pmatrix}J_B & 0\\ 0 & L_B\end{pmatrix}\right)\\&=\operatorname{rank}(\mu J_A-\lambda J_B)+\operatorname{rank}(\mu L_A-\lambda L_B).\end{aligned}$$

注意到

$$\operatorname{rank}(\mu L_A-\lambda L_B)=\text{const},\quad \forall(\lambda,\mu)\in \mathbf{G}_{1,2},$$

所以 $(\alpha,\beta)\in\lambda(A,B)$ 的必要与充分条件是 $(\alpha,\beta)\in\mathbf{G}_{1,2}$ 使得

$$\operatorname{rank}(\beta J_A-\alpha J_B)<\max_{(\lambda,\mu)\in\mathbf{G}_{1,2}}\operatorname{rank}(\mu J_A-\lambda J_B),$$

即

$$(\alpha,\beta)\in\lambda(J_A,J_B).$$

□

1.3 广义特征值问题的稳定性

当 $\{A, B\}$ 为奇异对时，$\{A, B\}$ 的特征值与特征向量一般并不连续依赖于 A 与 B 的元素的变化. 例如

$$A=\begin{pmatrix}2 & 0\\ 0 & 0\end{pmatrix},\quad B=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}.$$

利用非齐次坐标，显然有 $\lambda(A, B)=\{2\}$，并且任一向量 $x=(\xi_1,\xi_2)^T\in\mathbf{C}^2$，只要 $\xi_1\neq 0$，它就是 $\{A, B\}$ 属于特征值 2 的特征向量. 现考虑

$$A_\varepsilon=\begin{pmatrix}2 & \varepsilon\\ \varepsilon & 0\end{pmatrix},\quad B_\varepsilon=\begin{pmatrix}1 & \varepsilon\\ \varepsilon & 0\end{pmatrix},\quad 0<\varepsilon\ll 1.$$

易知 $\{A_\varepsilon,B_\varepsilon\}$ 是正则对，$\lambda(A_\varepsilon,B_\varepsilon)=\{1,1\}$，并且 $\{A_\varepsilon,B_\varepsilon\}$ 属于特征值 1 的特征向量为 $x=(0, 1)^T$. 于是我们看到，当 $\varepsilon\to+0$ 时，$A_\varepsilon\to A$，$B_\varepsilon\to B$，但 $\{A_\varepsilon,B_\varepsilon\}$ 的特征值是 1 不趋于 2，特征向量 $(0,1)^T$ 不趋于 $\mathbf{C}^2$ 中任一满足 $\xi_1\neq 0$ 的向量 $(\xi_1,\xi_2)^T$.

当 $\{A, B\}$ 为正则对时，$\{A, B\}$ 的特征向量也并不一定连续依赖于 A 与 B 的元素的变化，比如对于

$$A_\varepsilon=\begin{pmatrix}1 & 0 & 0\\ 0 & 1+\varepsilon & 1\\ 0 & 0 & 1+\varepsilon\end{pmatrix},\quad B_\varepsilon=\begin{pmatrix}1+\varepsilon & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},$$

$\{A_\varepsilon, B_\varepsilon\}$ 的特征向量在 $\varepsilon=0$ 处不连续. 但是正则对的特征值却是连续地依赖于矩阵元素的变化. 下面的定理 1.4 将说明这一点.

在叙述定理1.4之前,还须指出: 因为 $\beta Ax=\alpha Bx$ 的特征值 (α,β) 分布在 $G_{1,2}$ 上,所以为了测量特征值扰动的大小,将采用 $G_{1,2}$ 上的弦度量 $\rho((\alpha,\beta),(\gamma,\delta))$(见第二章(4.38)),其中$(\alpha,\beta)$, $(\gamma,\delta)\in G_{1,2}$.

定理 1.4. 设 $\{A,B\}$ 与 $\{C,D\}$ 皆为 n 阶正则对, $A=(\alpha_{ij})$ 与 $B=(\beta_{ij})$ 有分解式 (1.22). 又设 $C=(\gamma_{ij})$, $D=(\delta_{ij})$, $\lambda(C,D)=\{(\gamma_i,\delta_i)\}$, 并且无妨假设

$$|\gamma_i|^2+|\delta_i|^2=1,\quad i=1,\cdots,n.$$

令

$$p=|\det P|,\tag{1.23}$$

$$M=\max_{i,j}\{\sqrt{|\alpha_{ij}|^2+|\beta_{ij}|^2},\ \sqrt{|\gamma_{ij}|^2+|\delta_{ij}|^2}\},\tag{1.24}$$

$$\|(A,B)-(C,D)\|_F=\sqrt{\|A-C\|_F^2+\|B-D\|_F^2},\tag{1.25}$$

$$\Delta=n^{\frac{1}{n}}(n-1)^{1-\frac{1}{n}}M^{1-\frac{1}{n}}\|(A,B)-(C,D)\|_F^{\frac{1}{n}}\tag{1.26}$$

和

$$\Delta_1=\Delta/p^{\frac{1}{n}}.\tag{1.27}$$

则对于任一 $(\gamma,\delta)\in\lambda(C,D)$, 必有 $(\alpha_i,\beta_i)\in\lambda(A,B)$, 使得

$$\rho((\alpha_i,\beta_i),(\gamma,\delta))<\Delta_1;\tag{1.28}$$

并且当 Δ_1 足够小时,存在 $1,\cdots,n$ 的一个适当的排列 $\pi(1),\cdots,\pi(n)$, 使得

$$\rho((\alpha_i,\beta_i),(\gamma_{\pi(i)},\delta_{\pi(i)}))<(2n-1)\Delta_1.\tag{1.29}$$

证明:

1) 令

$$\varphi(\alpha,\beta)=\det(\beta A-\alpha B),\quad \phi(\alpha,\beta)=\det(\beta C-\alpha D),$$

$(\alpha,\beta)\in G_{1,2}, |\alpha|^2+|\beta|^2=1$.

首先可证

$$|\varphi(\alpha,\beta)-\psi(\alpha,\beta)|<\Delta^n. \tag{1.30}$$

事实上，有

$$\begin{aligned}
&|\varphi(\alpha,\beta)-\psi(\alpha,\beta)|\\
&\leqslant \sum_{(k,l)}\left|\det\begin{pmatrix}\beta\alpha_{11}-\alpha\beta_{11} & \cdots & \cdots & \beta\alpha_{1n}-\alpha\beta_{1n}\\ \cdots & \cdots & \cdots & \cdots\\ \beta\alpha_{k1}-\alpha\beta_{k1}\cdots\beta\alpha_{kl}-\alpha\beta_{kl} & & \beta\gamma_{k,l+1}-\alpha\delta_{k,l+1}\cdots\beta\gamma_{kn}-\alpha\delta_{kn} & \\ \cdots & \cdots & \cdots & \cdots\\ \beta\gamma_{n1}-\alpha\delta_{n1} & \cdots & \cdots & \beta\gamma_{nn}-\alpha\delta_{nn}\end{pmatrix}\right.\\
&\left.-\det\begin{pmatrix}\beta\alpha_{11}-\alpha\beta_{11} & \cdots & \cdots & \beta\alpha_{1n}-\alpha\beta_{1n}\\ \cdots & \cdots & \cdots & \cdots\\ \beta\alpha_{k1}-\alpha\beta_{k1}\cdots\beta\alpha_{k,l-1}-\alpha\beta_{k,l-1} & & \beta\gamma_{kl}-\alpha\delta_{kl}\cdots\beta\gamma_{kn}-\alpha\delta_{kn} & \\ \cdots & \cdots & \cdots & \cdots\\ \beta\gamma_{n1}-\alpha\delta_{n1}\cdots & \cdots & \cdots & \beta\gamma_{nn}-\alpha\delta_{nn}\end{pmatrix}\right|\\
&=\sum_{(k,l)}|(\beta\alpha_{kl}-\alpha\beta_{kl})-(\beta\gamma_{kl}-\alpha\delta_{kl})||\det T_{kl}|.
\end{aligned}$$

其中 $\sum\limits_{(k,l)}$ 表示对指标 (k,l) 按照自然顺序

$$(1.1),\cdots,(1,n),(2,1),\cdots,(2,n),\cdots,(n,1),\cdots,(n,n)$$

求和；$T_{kl}\in C^{(n-1)\times(n-1)}$，它的每一行的 Euclid 向量范数不超过 $(n-1)M$，因此根据 Hadamard 不等式，

$$|\det T_{kl}|\leqslant[(n-1)M]^{n-1};$$

此外，

$$\begin{aligned}&|(\beta\alpha_{kl}-\alpha\beta_{kl})-(\beta\gamma_{kl}-\alpha\delta_{kl})|\\ &\qquad\leqslant\sqrt{|\alpha_{kl}-\gamma_{kl}|^2+|\beta_{kl}-\delta_{kl}|^2}\equiv\varepsilon_{kl}.\end{aligned}$$

所以

$$\begin{aligned}|\varphi(\alpha,\beta)-\psi(\alpha,\beta)|&\leqslant\sum_{(k,l)}\varepsilon_{kl}[(n-1)M]^{n-1}\\ &\leqslant n\sqrt{\sum_{(k,l)}\varepsilon_{kl}^2}\,[(n-1)M]^{n-1}\\ &=\Delta^n.\end{aligned}$$

2) 设 (γ,δ) 是 $\{C,D\}$ 的任一特征值，并且 $|\gamma|^2+|\delta|^2=1$，则由 (1.23) 和 (1.30) 可知

$$\prod_{i=1}^n|\delta\alpha_i-\gamma\beta_i|=\frac{1}{p}|\varphi(\gamma,\delta)|$$

$$= \frac{1}{p} |\varphi(\gamma,\delta) - \psi(\gamma,\delta)| < \Delta^n/p.$$

注意到上式左端等于 $\prod_{i=1}^{n} \rho((\alpha_i,\beta_i),(\gamma,\delta))$，所以，有

$$\prod_{i=1}^{n} \rho((\alpha_i,\beta_i),(\gamma,\delta)) < \frac{\Delta^n}{p} = \Delta_1^n;$$

因此，至少有一个 (α_i,β_i)，适合

$$\rho((\alpha_i,\beta_i),(\gamma,\delta)) < \Delta_1.$$

3) 令

$$\mathscr{D}_{\Delta_1}(\alpha_i,\beta_i) = \{(\alpha,\beta) \in \mathrm{G}_{1,2}: \rho((\alpha,\beta),(\alpha_i,\beta_i)) \leqslant \Delta_1\},$$
$$i = 1,\cdots,n. \tag{1.31}$$

并假设 Δ_1 足够小，使得

$$\mathrm{G}_{1,2} \Big\backslash \bigcup_{i=1}^{n} \mathscr{D}_{\Delta_1}(\alpha_i, \beta_i) \neq \emptyset. \tag{1.32}$$

如果 $\rho((\alpha_i,\beta_i), (\alpha_j,\beta_j)) < 2\Delta_1$，则称 (α_i, β_i) 与 (α_j, β_j) 相邻；如果存在 $\mathrm{G}_{1,2}$ 上的点列 $(\alpha_{\nu_1}, \beta_{\nu_1}), \cdots, (\alpha_{\nu_k}, \beta_{\nu_k})$，使得 $(\alpha_i,\beta_i),(\alpha_{\nu_1},\beta_{\nu_1}),\cdots,(\alpha_{\nu_k},\beta_{\nu_k}),(\alpha_j,\beta_j)$ 中，除 (α_i,β_i) 外，每一点都与前面的一点相邻，则称 (α_i, β_i) 与 (α_j, β_j) 相连．现将点集 $\{(\alpha_i,\beta_i)\}_{i=1}^{n}$ 划分成 s 个子集 $\Omega_1, \cdots, \Omega_s$，划分的原则是：所有相连的点属于同一个子集，不相连的点属于不同的子集．

令

$$G_k = \bigcup_{(\alpha_i,\beta_i) \in \Omega_k} \mathscr{D}_{\Delta_1}(\alpha_i,\beta_i), k = 1,\cdots,s. \tag{1.33}$$

G_k 的边界记作 ∂G_k，它由有穷多个非 Euclid 圆弧构成．

引进函数

$$\left.\begin{array}{l} \chi_t(\alpha,\beta) = \varphi(\alpha,\beta) + t(\psi(\alpha,\beta) - \varphi(\alpha,\beta)), \\ (\alpha,\beta) \in \mathrm{G}_{1,2},\ |\alpha|^2 + |\beta|^2 = 1,\ 0 \leqslant t \leqslant 1. \end{array}\right\} \tag{1.34}$$

可证

$$\chi_t(\alpha,\beta) \neq 0,\quad \forall(\alpha,\beta) \in \bigcup_{k=1}^{s} \partial G_k,\ \forall t \in [0,1]. \tag{1.35}$$

事实上，如果对于某一个 $t\in[0,1]$，存在

$$(\alpha',\beta'),\ |\alpha'|^2+|\beta'|^2=1,$$

使得 $\chi_t(\alpha',\beta')=0$，则由

$$p\prod_{i=1}^{n}|\beta'\alpha_i-\alpha'\beta_i|=|\varphi(\alpha',\beta')|$$
$$=t|\psi(\alpha',\beta')-\varphi(\alpha',\beta')|<\Delta^n$$

即

$$\prod_{i=1}^{n}|\beta'\alpha_i-\alpha'\beta_i|<\frac{\Delta^n}{p}=\Delta_1^n$$

可知，必有某一 $(\alpha_{i'},\beta_{i'})$，使得

$$\rho((\alpha',\beta'),(\alpha_{i'},\beta_{i'}))=|\beta'\alpha_{i'}-\alpha'\beta_{i'}|<\Delta_1,$$

即 (α',β') 必位于某一个 G_k 内。所以，我们的断言(1.35)成立。

4）根据(1.32)，在 C 内存在点 $z_0\bar{\in}\bigcup\limits_{k=1}^{s}G_k$。令

$$\frac{1}{\sqrt{1+|z_0|^2}}=\sigma,\quad \frac{z_0}{\sqrt{1+|z_0|^2}}=\tau,$$

作 Möbius 变换

$$\begin{pmatrix}\xi\\ \eta\end{pmatrix}=\begin{pmatrix}\bar{\tau} & \bar{\sigma}\\ \sigma & -\tau\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix},$$

即

$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix}=\begin{pmatrix}\tau & \bar{\sigma}\\ \sigma & -\bar{\tau}\end{pmatrix}\begin{pmatrix}\xi\\ \eta\end{pmatrix}\tag{1.36}$$

（它是把点 z_0 变到无穷远点∞的分式线性变换）。于是

$$\varphi(\alpha,\beta)\equiv\tilde{\varphi}(\xi,\eta),\ \psi(\alpha,\beta)\equiv\tilde{\psi}(\xi,\eta),$$
$$\chi_t(\alpha,\beta)\equiv\tilde{\chi}_t(\xi,\eta);\tag{1.37}$$

同时，$G_1,\cdots,G_s$ 分别变为 $\tilde{G}_1,\cdots,\tilde{G}_s$，并且

$$\bigcup_{k=1}^{s}\tilde{G}_k\subset C.\tag{1.38}$$

今在 $\bigcup\limits_{k=1}^{s}\tilde{G}_k$ 上考虑函数

$$\tilde{\chi}_t(\xi,\eta) = \tilde{\varphi}(\xi,\eta) + t(\tilde{\phi}(\xi,\eta) - \tilde{\varphi}(\xi,\eta)).$$

由 (1.38) 知

$$z = \frac{\xi}{\eta} \in \mathbf{C}, \quad \forall(\xi,\eta) \in \bigcup_{k=1}^{s} \widetilde{G}_k.$$

因此，可记

$$\tilde{\varphi}(\xi,\eta) \equiv \tilde{\varphi}(z), \quad \tilde{\phi}(\xi,\eta) \equiv \tilde{\phi}(z),$$

$$\tilde{\chi}_t(\xi,\eta) \equiv \tilde{\chi}_t(z), \quad z \in \bigcup_{k=1}^{s} \widetilde{G}_k.$$

它们适合关系式

$$\tilde{\chi}_t(z) = \tilde{\varphi}(z) + t(\tilde{\phi}(z) - \tilde{\varphi}(z)). \tag{1.39}$$

根据 (1.35)，

$$\tilde{\chi}_t(z) \neq 0, \quad \forall z \in \bigcup_{k=1}^{s} \partial\widetilde{G}_k, \forall t \in [0,1].$$

所以，对于任意固定的 $t \in [0, 1]$ 和任意固定的 $k(1 \leqslant k \leqslant s)$，有

$$p_k \equiv \min_{z \in \partial\widetilde{G}_k} |\tilde{\chi}_t(z)| > 0.$$

记

$$\zeta(z) = \omega(\tilde{\phi}(z) - \tilde{\varphi}(z)), \ \omega \in \mathbf{R}.$$

取绝对值充分小的 ω，使得

$$\max_{z \in \partial\widetilde{G}_k} |\zeta(z)| < p_k;$$

于是 $\widetilde{G}_k$ 上的解析函数 $\tilde{\chi}_t(z)$ 与 $\zeta(z)$ 在 $\partial\widetilde{G}_k$ 上恒满足

$$|\tilde{\chi}_t(z)| > |\zeta(z)|.$$

据 Rouché 定理(见第三章 § 1 的脚注)，

$$\tilde{\chi}_t(z) + \zeta(z) = \tilde{\varphi}(z) + (t + \omega)(\tilde{\phi}(z) - \tilde{\varphi}(z))$$

与

$$\tilde{\chi}_t(z) = \tilde{\varphi}(z) + t(\tilde{\phi}(z) - \tilde{\varphi}(z))$$

在 $\widetilde{G}_k$ 内有相同个数的零点. 因此，在 $[0,1]$ 上任一点 t 的邻域内，$\tilde{\chi}_t(z)$ 的零点个数 $N(t)$ 是个常数. 从而可知 $N(t)$ 是 t 在 $[0, 1]$ 上的连续函数. 注意到 $N(t)$ 只可能是自然数，所以

$N(t)=$ 常数 ($t\in[0,\ 1]$). 特别地，在(1.39)中分别取 $t=0$ 与 $t=1$，得到 $\tilde{\varphi}(z)$ 与 $\tilde{\phi}(z)$，它们在每个 $\tilde{G}_k$ $(k=1,\cdots,s)$ 内有相同个数的零点. 再利用变换 (1.36) 可知，$\varphi(\alpha,\beta)=0$ 与 $\phi(\alpha,\beta)=0$ 在每个 $G_k(k=1,\cdots,s)$ 内有相同个数的零点.

考虑任一固定的 $k(1\leqslant k\leqslant s)$. 假定 $\varphi(\alpha,\beta)=0$ 与 $\phi(\alpha,\beta)=0$ 在 G_k 内各有 N_k 个根. 根据 G_k 的构造可知,在 G_k 内, $\varphi(\alpha,\beta)=0$ 的每个根 (α_i,β_i) 与 $\phi(\alpha,\beta)=0$ 的每个根 (γ_j,δ_j) 的距离适合

$$\rho((\alpha_i,\beta_i),(\gamma_j,\delta_j))\leqslant(2N_k-1)\Delta_1\leqslant(2n-1)\Delta_1.$$

所以存在 $\{(\gamma_j,\delta_j)\}$ 的一个适当的排列 $(\gamma_{\pi(1)},\delta_{\pi(1)}),\cdots,(\gamma_{\pi(n)},\delta_{\pi(n)})$，使得

$$\rho((\alpha_i,\beta_i),(\gamma_{\pi(i)},\delta_{\pi(i)}))\leqslant(2n-1)\Delta_1.$$ □

注 1.1 所谓广义特征值问题的稳定性，指的是矩阵对 $\{A,B\}$ 的特征值与特征向量是否连续地依赖于 A 与 B 的元素的变化. 就这个意义来说，定理 1.4 表明正则对的特征值是稳定的,即对于任一固定的正则对 $\{A,B\}$ 以及不断变化的正则对 $\{C,D\}$, 如果矩阵 C 与 D 分别趋于 A 与 B，则 $\{C,D\}$ 的广义特征值趋于 $\{A,B\}$ 的广义特征值. 但这并不等于说,对于任意两个正则对 $\{A,B\}$ 与 $\{C,D\}$，只要 C 与 D 分别靠近 A 与 B，$\{C,D\}$ 的广义特征值就一定靠近 $\{A,B\}$ 的广义特征值. 这一点值得注意. 例如，考察下列两个正则对

$$\{A,B\}=\left\{\begin{pmatrix}\frac{1}{\sqrt{2}} & 0\\ 0 & 0\end{pmatrix},\begin{pmatrix}\frac{1}{\sqrt{2}} & 0\\ 0 & \varepsilon\end{pmatrix}\right\},$$

$$\{C,D\}=\left\{\begin{pmatrix}\frac{1}{\sqrt{2}} & 0\\ 0 & \varepsilon\end{pmatrix},\begin{pmatrix}\frac{1}{\sqrt{2}} & 0\\ 0 & 0\end{pmatrix}\right\},\ 0<\varepsilon\ll 1.$$

显然

$$\lambda(A,B)=\{(1,1),(0,1)\},$$
$$\lambda(C,D)=\{(1,1),(1,0)\},$$

利用非齐次坐标，有

$$\lambda(A,B)=\{1,0\},\quad \lambda(C,D)=\{1,\infty\}.$$

$\|C-A\|_2=\|D-B\|_2=\varepsilon\ll 1$ 表明 C 与 D 分别靠近 A 与 B，但是 $\{C,D\}$ 的特征值(1,0)(其非齐次坐标为∞)并不靠近 $\{A,B\}$ 的任何特征值. 如果应用定理 1.4，则 (1.23)—(1.27) 诸式所示的量，分别为

$$p=\left|\det\begin{pmatrix}1 & 0\\ 0 & \varepsilon\end{pmatrix}\right|=\varepsilon,\ M=1,$$

$$\|(A,B)-(C,D)\|_F=\sqrt{2}\,\varepsilon,$$

$$\Delta=8^{\frac{1}{4}}\sqrt{\varepsilon},\quad \Delta_1=8^{\frac{1}{4}}.$$

代入 (1.29) 式也可以看出，无论正数 ε 如何小，根据定理 1.4 所得到的特征值的扰动界限并不小.

设 $\{A,B\}$ 与 $\{C,D\}$ 均为 n 阶正则对，

$$\lambda(A,B)=\{(\alpha_i,\beta_i)\},\ \lambda(C,D)=\{(\gamma_i,\delta_i)\}.$$

为了表示 $\lambda(A,B)$ 与 $\lambda(C,D)$ 之间的差距，特地引进下述几个量:

1) $\{C,D\}$ 对 $\{A,B\}$ 的谱改变量

$$s_{\{A,B\}}\{C,D\}\equiv\max_{1\leqslant i\leqslant n}\min_{1\leqslant j\leqslant n}\rho((\alpha_i,\beta_i),(\gamma_j,\delta_j));\qquad(1.40)$$

2) $\{A,B\}$ 与 $\{C,D\}$ 的特征值改变量

$$v(\{A,B\},\{C,D\})\equiv\min_{\pi}\max_{1\leqslant i\leqslant n}\rho((\alpha_i,\beta_i),(\gamma_{\pi(i)},\delta_{\pi(i)})),\qquad(1.41)$$

其中 π 是 $1,\cdots,n$ 的某一个排列，上式右端的 $\min\limits_{\pi}$ 表示在 $1,\cdots,n$ 的所有可能的排列上取最小值;

3) $\{A,B\}$ 与 $\{C,D\}$ 的谱的 Euclid 距离

$$e(\{A,B\},\{C,D\})\equiv\min_{\pi}\sqrt{\sum_{i=1}^{n}\rho^2((\alpha_i,\beta_i),(\gamma_{\pi(i)},\delta_{\pi(i)}))},\qquad(1.42)$$

其中 π 是 $1,\cdots,n$ 的某个排列，上式右端的 $\min\limits_{\pi}$ 表示在 $1,\cdots,n$ 的所有可能的排列上取最小值.

利用上述记号，定理 1.4 的结论 (1.28) 和 (1.29) 可以简单地记作

$$s_{\{A,B\}}\{C,D\} \leqslant \Delta_1,$$
$$v(\{A,B\},\{C,D\}) < (2n-1)\Delta_1.$$

1.4 几类重要的正则对

定义 1.4. 设$\{A,B\}$是n阶正则对．如果$\mathbf{C}^n$存在由$\{A,B\}$的特征向量所构成的一组基底，则称$\{A,B\}$为可对角化对（或可正规化对）；如果$\mathbf{C}^n$存在由$\{A, B\}$的特征向量所构成的一组标准正交基．则称$\{A,B\}$为正规对．

所有n阶可对角化对的全体，记作$\mathrm{D_g}(n)$；所有n阶正规对的全体，记作$\mathrm{N}(n)$．

定理 1.5[87]．设$\{A, B\}$是n阶正则对．则$\{A, B\} \in \mathrm{D_g}(n)$的必要与充分条件是存在非奇异阵$S,Q \in \mathbf{C}^{n\times n}$，使得

$$S^HAQ = \mathrm{diag}(\alpha_1,\cdots,\alpha_n), S^HBQ = \mathrm{diag}(\beta_1,\cdots,\beta_n). \quad (1.43)$$

证明:

充分性．设A, B有分解式(1.43)，记$Q=(\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n)$，$\boldsymbol{q}_i \in \mathbf{C}^n, i=1,\cdots,n$．则由$\{A, B\}$是正则对，以及存在分解式(1.43)可知

$$(\alpha_i,\beta_i) \neq (0,0), \quad \boldsymbol{q}_i \neq 0,$$
$$\beta_iA\boldsymbol{q}_i = \alpha_iB\boldsymbol{q}_i, \quad i=1,\cdots,n.$$

即$\boldsymbol{q}_i$是$\{A,B\}$属于特征值(α_i, β_i)的特征向量，$i=1,\cdots,n$．因此，$\mathbf{C}^n$存在由$\{A,B\}$的特征向量组成的一组基底$\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n$．所以$\{A,B\}$是可对角化对．

必要性．设$\{A,B\}$存在n个线性无关的特征向量$\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n$:

$$\beta_iA\boldsymbol{q}_i = \alpha_iB\boldsymbol{q}_i, \quad (\alpha_i,\beta_i) \neq (0,0), \quad i=1,\cdots,n.$$

则可根据$\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n$构造$n$个向量$\boldsymbol{r}_1,\cdots,\boldsymbol{r}_n$:

$$\boldsymbol{r}_i = \begin{cases} A\boldsymbol{q}_i/\alpha_i & \text{当 } \alpha_i \neq 0 \text{ 时} \\ B\boldsymbol{q}_i/\beta_i & \text{当 } \beta_i \neq 0 \text{ 时}, \end{cases} \quad i=1,\cdots,n. \quad (1.44)$$

令

$$Q=(\boldsymbol{q}_1,\cdots,\boldsymbol{q}_n), \; R=(\boldsymbol{r}_1,\cdots,\boldsymbol{r}_n);$$

$$\Lambda = \mathrm{diag}(\alpha_1, \cdots, \alpha_n), \quad \Omega = \mathrm{diag}(\beta_1, \cdots, \beta_n),$$

则(1.44)表示

$$\Lambda Q = R\Lambda, \quad BQ = R\Omega.$$

于是

$$A = R\Lambda Q^{-1}, \quad B = R\Omega Q^{-1}. \tag{1.45}$$

再根据 $\{A, B\}$ 是正则对，存在 $\lambda \in \mathbb{C}$，使得 $\det(A + \lambda B) \neq 0$，即

$$\det R \cdot \det(\Lambda + \lambda\Omega) \det Q^{-1} \neq 0.$$

所以 R 必为非奇异阵. 令 $S^H = R^{-1}$，则由 (1.45) 立即得到分解式 (1.43). □

同理可证

定理 1.6. 设 $\{A, B\}$ 是 n 阶正则对. 则 $\{A, B\} \in \mathrm{N}(n)$ 的必要与充分条件是存在非奇异阵 S 和酉阵 U，使得

$$S^H AU = \mathrm{diag}(\alpha_1, \cdots, \alpha_n), S^H BU = \mathrm{diag}(\beta_1, \cdots, \beta_n). \tag{1.46}$$

定义 1.5. 设 $A, B \in \mathbb{C}^{n \times n}$ 均为 Hermite 阵. 如果

$$c(A, B) \equiv \min_{\|x\|_2 = 1} |x^H(A + iB)x| > 0, \tag{1.47}$$

则称 $\{A, B\}$ 为 n 阶定型对，$c(A, B)$ 叫做定型对 $\{A, B\}$ 的 Crawford 数.

所有 n 阶定型对的全体，记作 $\mathrm{D}(n)$.

定理1.7.[159] 设 $\{A, B\} \in \mathrm{D}(n)$. 令

$$A_\varphi = A\cos\varphi - B\sin\varphi, \tag{1.48}$$

$$B_\varphi = A\sin\varphi + B\cos\varphi, \ \varphi \in \mathbb{R}.$$

则存在 $\varphi \in [0, 2\pi)$，使得 (1.48) 式中的 B_φ 为正定阵，并且 $\{A, B\}$ 的 Crawford 数

$$c(A, B) = \lambda_{\min}(B_\varphi), \tag{1.49}$$

此处 $\lambda_{\min}(B_\varphi)$ 表示 B_φ 的最小特征值.

证明:

首先考虑 $A + iB$ 的值域

$$F(A + iB) \equiv \{x^H(A + iB)x : x \in \mathbb{C}^n, \|x\|_2 = 1\}.$$

根据第一章 §4，$F(A + iB)$ 是 $\mathbb{C}$ 内一有界闭凸集；再由 (1.47)

可知，$F(A+iB)$ 不包含原点.

设(1.47)中的最小值在 $\boldsymbol{x}_0\in\mathbb{C}^n$ 达到. 记

$$\zeta_0=\boldsymbol{x}_0^H(A+iB)\boldsymbol{x}_0,$$

则有

$$|\zeta_0|=\min\{|\zeta|:\zeta\in F(A+iB)\}=c(A,B)>0.$$

因此 $F(A+iB)$ 必位于不包含原点的闭半平面

$$\mathscr{H}=\{z\in\mathbb{C}:\mathrm{Re}(\bar{\zeta}_0 z)\geqslant|\zeta_0|^2\}$$

之中(如图 4-1 所示).

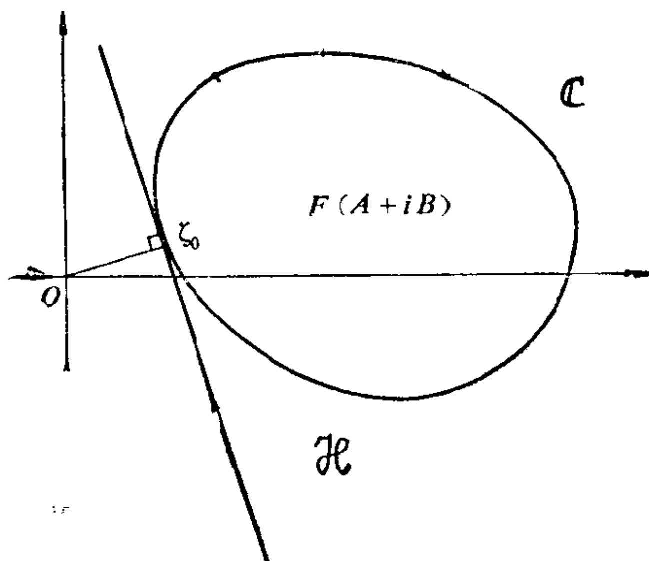


图 4-1

注意到，(1.48) 式可写为

$$A_\varphi+iB_\varphi=e^{i\varphi}(A+iB). \tag{1.50}$$

并且易知,经变换 (1.50)，$c(A_\varphi,B_\varphi)=c(A,B)$，同时，$F(A+iB)$ 与 $\mathscr{H}$ 按反时针方向旋转 φ 角，分别变为 $F(A_\varphi+iB_\varphi)$ 和 $\mathscr{H}_\varphi$. 取 φ 使得 $\mathscr{H}_\varphi$ 位于上半平面之中,则 $\boldsymbol{x}_0^H A_\varphi\boldsymbol{x}_0=0$. 这时，对于 $\mathbb{C}^n$ 中的任一单位向量 $\boldsymbol{x}$，有

$$\boldsymbol{x}^H B_\varphi\boldsymbol{x}\geqslant c(A_\varphi,B_\varphi)=c(A,B)>0.$$

因此，

$$\min_{\|\boldsymbol{x}\|_2=1}\boldsymbol{x}^H B_\varphi\boldsymbol{x}=\boldsymbol{x}_0^H B_\varphi\boldsymbol{x}_0=c(A,B)>0. \tag{1.51}$$

(1.51) 式表明，$B_\varphi>0$，$\boldsymbol{x}_0$ 是 B_φ 对应于 $\lambda_{\min}(B_\varphi)$ 的特征向量(本节习题 1)，并且 $c(A,B)=\lambda_{\min}(B_\varphi)$. □

容易证明，对于(1.48)式中的 Hermite 阵 A_φ 与 B_φ，当 B_φ

为正定阵时，必存在非奇异阵 Q，使得 $Q^H A_\varphi Q$ 与 $Q^H B_\varphi Q$ 同时为对角阵（本节习题 2）；因而，对于（1.48）式中与 A_φ 和 B_φ 相对应的 A 与 B，$Q^H A Q$ 与 $Q^H B Q$ 同时为对角阵．这就证明了

定理 1.8. 设 $\{A, B\} \in \mathbb{D}(n)$．则存在非奇异阵 $Q \in \mathbb{C}^{n\times n}$，使得

$$Q^H A Q = \operatorname{diag}(\alpha_1, \cdots, \alpha_n),$$
$$Q^H B Q = \operatorname{diag}(\beta_1, \cdots, \beta_n). \tag{1.52}$$

进而可证

推论 1.6. 任一定型对 $\{A, B\}$ 必为正则对．

证明:

根据定义 1.5，

$$|\boldsymbol{x}^H A \boldsymbol{x}|^2 + |\boldsymbol{x}^H B \boldsymbol{x}|^2 \geqslant c^2(A, B) > 0,$$
$$\forall \boldsymbol{x} \in \mathbb{C}^n, \quad \|\boldsymbol{x}\|_2 = 1.$$

利用分解式（1.52）中的 Q，令 $Q^{-1}\boldsymbol{x} = \boldsymbol{y}$，则有

$$|\boldsymbol{y}^H \operatorname{diag}(\alpha_1, \cdots, \alpha_n)\boldsymbol{y}|^2 + |\boldsymbol{y}^H \operatorname{diag}(\beta_1, \cdots, \beta_n)\boldsymbol{y}|^2 > 0,$$
$$\forall \boldsymbol{y} \in \mathbb{C}^n, \quad \boldsymbol{y} \neq 0.$$

由此可知 $(\alpha_i, \beta_i) \neq (0, 0)$，$i = 1, \cdots, n$．因而

$$\det(A + \lambda B) = |\det Q|^{-2} \prod_{i=1}^{n} (\alpha_i + \lambda\beta_i) \not\equiv 0, \lambda \in \mathbb{C}.$$

所以 $\{A, B\}$ 是正则对．□

定理 1.5，定理 1.6 和定理 1.8 表明，可对角化对、正规对和定型对分别是可对角化阵、正规阵和 Hermite 阵的推广，并且

$$\mathbb{N}(n), \mathbb{D}(n) \subset \mathbb{D}_g(n).$$

值得指出的是，Hermite 阵必是正规阵，但定型对不一定是正规对．

习题

1. 设 $H \in \mathbb{C}^{n\times n}$ 为 Hermite 阵．$\boldsymbol{x}_0 \in \mathbb{C}^n$，$\|\boldsymbol{x}_0\|_2 = 1$，满足

$$\min_{\|\boldsymbol{x}\|_2=1} \boldsymbol{x}^H H \boldsymbol{x} = \boldsymbol{x}_0^H H \boldsymbol{x}_0 = \alpha.$$

证明 $\boldsymbol{x}_0$ 是 H 属于特征值 α 的特征向量．

2. 设 $A,B\in\mathbb{C}^{n\times n}$ 均为 Hermite 阵，并且 $B>0$. 试证：存在非奇异阵 $X=(x_1,\cdots,x_n)\in\mathbb{C}^{n\times n}$，使得

$$X^HAX=\mathrm{diag}(\gamma_1,\cdots,\gamma_n),\quad X^HBX=I,$$

并且 $Ax_i=\gamma_iBx_i, i=1,\cdots,n$.

3. 设 $\{A,B\}\in\mathbb{D}(n)$，其中

$$B=\begin{pmatrix} I^{(r)} & 0 \\ 0 & 0 \end{pmatrix},\quad 1\leqslant r\leqslant n-1.$$

如果将 A 分块为 $A=\begin{pmatrix} A_{11} & A_{12} \\ A_{12}^H & A_{22} \end{pmatrix}$，其中 $A_{11}\in\mathbb{C}^{r\times r}$，则必有 $A_{22}>0$ 或者 $A_{22}<0$.

4. 设 $A,B\in\mathbb{C}^{n\times n}$. 试证：存在酉阵 U 与 V，使得

$$A=U\mathrm{diag}(\alpha_1,\cdots,\alpha_n)V,$$
$$B=U\mathrm{diag}(\beta_1,\cdots,\beta_n)V$$

的必要与充分条件是 AB^H 与 B^HA 均为正规阵。

§2 Gerschgorin 理论

2.1 Gerschgorin 型定理

首先证明一个一般性的结论.

定理 2.1[22]. 设 $\{A,B\}$ 与 $\{C,D\}$ 均为 n 阶正则对. 令

$$\mathscr{D}_{\{A,B\}}=\{(\alpha,\beta)\in\mathbb{G}_{1,2}:\|(\beta C-\alpha D)^{-1}[\beta(C-A)-\alpha(D-B)]\| \geqslant 1 \text{ 或 } \det(\beta C-\alpha D)=0\},\tag{2.1}$$

其中$\|\ \ \|$表示任一种相容的矩阵范数. 则

$$\lambda(A,B)\subset\mathscr{D}_{\{A,B\}}.\tag{2.2}$$

证明:

任取 $(\alpha,\beta)\in\lambda(A,B)$，无妨假设 $|\alpha|^2+|\beta|^2=1$. 有两种可能:

1) $(\alpha,\beta)\in\lambda(C,D)$，即有 $\det(\beta C-\alpha D)=0$.

2) $(\alpha,\beta)\notin\lambda(C,D)$，即 $\beta C-\alpha D$ 可逆. 设 x 是 $\{A,B\}$ 属于 (α,β) 的特征向量，则

$$\beta C\boldsymbol{x} - \alpha D\boldsymbol{x} = [\beta(C - A) - \alpha(D - B)]\boldsymbol{x}.$$

由此可知

$$\|(\beta C - \alpha D)^{-1}[\beta(C - A) - \alpha(D - B)]\| \geqslant 1. \qquad \square$$

注意到，对于 $A = (\alpha_{ij}), B = (\beta_{ij}) \in \mathbb{C}^{n\times n}$，如果 $\{A, B\}$ 是正则对，则必存在排列方阵 P，使得 $AP = (\alpha'_{ij})$ 和 $BP = (\beta'_{ij})$ 满足 $(\alpha'_{ii}, \beta'_{ii}) \neq (0,0), i = 1, \cdots, n$. 所以，在本节内，无妨假设正则对 $\{A, B\}$ 满足

$$(\alpha_{ii}, \beta_{ii}) \neq (0,0), \quad i = 1, \cdots, n. \tag{2.3}$$

定理 2.2.[22] 设 $\{A, B\}$ 为 n 阶正则对，满足条件 (2.3). 令

$$D_i(A, B) = \left\{(\alpha, \beta) \in \mathbb{G}_{1,2}: |\beta\alpha_{ii} - \alpha\beta_{ii}| \leqslant \sum_{j\neq i} |\beta\alpha_{ij} - \alpha\beta_{ij}|\right\}, \quad i = 1, \cdots, n. \tag{2.4}$$

则

$$\lambda(A, B) \subset \bigcup_{i=1}^{n} D_i(A, B).$$

证明:

在定理 2.1 中，取

$$C = \mathrm{diag}(\alpha_{11}, \cdots, \alpha_{nn}) \text{ 和 } D = \mathrm{diag}(\beta_{11}, \cdots, \beta_{nn}),$$

并取范数 $\| \ \|_\infty$，则 (2.1) 为

$$\mathscr{D}_{\{A,B\}} = \left\{(\alpha, \beta) \in \mathbb{G}_{1,2}: \max_i \frac{\sum_{j\neq i} |\beta\alpha_{ij} - \alpha\beta_{ij}|}{|\beta\alpha_{ii} - \alpha\beta_{ii}|} \geqslant 1 \text{ 或 } \prod_{i=1}^{n} (\beta\alpha_{ii} - \alpha\beta_{ii}) = 0 \right\}.$$

容易验证，

$$\mathscr{D}_{\{A,B\}} = \bigcup_{i=1}^{n} D_i(A, B).$$

并利用 (2.2)，立即得到

$$\lambda(A, B) \subset \bigcup_{i=1}^{n} D_i(A, B). \qquad \square$$

注 2.1. 可以把(2.4)式所示的 $D_i(A,B)(i=1,\cdots,n)$ 称为 $\{A,B\}$ 的 Gerschgorin 域. 如果在定理 2.2 中取 $B=I^{(n)}$, 则 $D_i(A,B)$ 便是 A 的 Gerschgorin 圆盘 $G_i(A), i=1,\cdots,n$ (见第三章§2(2.3)).

对于满足条件 (2.3) 的 $A=(\alpha_{ij})$ 与 $B=(\beta_{ij})\in\mathbb{C}^{n\times n}$, 令

$$\boldsymbol{a}_i=(\alpha_{i1},\cdots,\alpha_{i,i-1},\alpha_{i,i+1},\cdots,\alpha_{in})^T,$$

$$\boldsymbol{b}_i=(\beta_{i1},\cdots,\beta_{i,i-1},\beta_{i,i+1},\cdots,\beta_{in})^T.$$

利用弦度量, (2.4)式可改写为

$$D_i(A,B)=\left\{(\alpha,\beta)\in\mathbb{G}_{1,2}:\ \rho((\alpha_{ii},\beta_{ii}),(\alpha,\beta))\leqslant\frac{\|\alpha\boldsymbol{b}_i-\beta\boldsymbol{a}_i\|_1}{\sqrt{(|\alpha|^2+|\beta|^2)(|\alpha_{ii}|^2+|\beta_{ii}|^2)}}\right\}$$

$$i=1,\cdots,n. \tag{2.5}$$

以下用 $\mathscr{D}_\chi(\alpha_1,\beta_1)$ 表示以 $(\alpha_1,\beta_1)\in\mathbb{G}_{1,2}$ 为中心, 以 $\chi\geqslant 0$ 为半径的非 Euclid 圆盘, 即

$$\mathscr{D}_\chi(\alpha_1,\beta_1)=\{(\alpha,\beta)\in\mathbb{G}_{1,2}:\rho((\alpha_1,\beta_1),(\alpha,\beta))\leqslant\chi\}. \tag{2.6}$$

注意到(2.5)式右端括号内的

$$\begin{aligned}\|\alpha\boldsymbol{b}_i-\beta\boldsymbol{a}_i\|_1&\leqslant\sum_{j\neq i}(|\alpha||\beta_{ij}|+|\beta||\alpha_{ij}|)\\&=\left\{(|\alpha|^2+|\beta|^2)\left[\left(\sum_{j\neq i}|\alpha_{ij}|\right)^2+\left(\sum_{j\neq i}|\beta_{ij}|\right)^2\right]\right.\\&\qquad\left.-\left[\sum_{j\neq i}(|\alpha||\alpha_{ij}|-|\beta||\beta_{ij}|)\right]^2\right\}^{\frac{1}{2}}\\&\leqslant\sqrt{|\alpha|^2+|\beta|^2}\sqrt{\|\boldsymbol{a}_i\|_1^2+\|\boldsymbol{b}_i\|_1^2},\end{aligned} \tag{2.7}$$

所以,如果令

$$G_i(A,B)=\mathscr{D}_{\chi_i}(\alpha_{ii},\beta_{ii}),$$

$$\chi_i=\sqrt{\frac{\|\boldsymbol{a}_i\|_1^2+\|\boldsymbol{b}_i\|_1^2}{|\alpha_{ii}|^2+|\beta_{ii}|^2}},\quad i=1,\cdots,n, \tag{2.8}$$

则据定理 2.2 和(2.5)—(2.8), 并利用定理 1.4, 采取与第三章定理2.2同样的证明方法,可导出正则对的 Gerschgorin 型定理如下:

定理 2.3[22]. 设 $\{A, B\}$ 是满足条件 (2.3) 的 n 阶正则对. 则

1)
$$\lambda(A,B) \subset \bigcup_{i=1}^{n} G_i(A,B), \tag{2.9}$$

其中 $G_i(A,B)$ 如(2.8)式所示,叫做 $\{A,B\}$ 的 Gerschgorin 非 Euclid 圆盘,(简称 G 圆盘) $i=1,\cdots,n$;

2) 如果 $\{A,B\}$ 的 $m(m<n)$ 个 G 圆盘的并集不与其它的 G 圆盘相交,则 $\{A,B\}$ 恰有 m 个特征值在此并集之中.

注 2.2. 令

$$\widetilde{G}_i(A,B) = \mathscr{D}_{\tilde{\chi}_i}(\alpha_{ii},\beta_{ii}),$$

$$\tilde{\chi}_i = \frac{\sum_{j\neq i}\sqrt{|\alpha_{ij}|^2+|\beta_{ij}|^2}}{\sqrt{|\alpha_{ii}|^2+|\beta_{ii}|^2}}, i=1,\cdots,n. \tag{2.10}$$

则由不等式

$$\sqrt{\|\boldsymbol{a}_i\|_1^2+\|\boldsymbol{b}_i\|_1^2} \leqslant \sum_{j\neq i}\sqrt{|\alpha_{ij}|^2+|\beta_{ij}|^2}$$

可知

$$G_i(A,B) \subseteq \widetilde{G}_i(A,B), \ i=1,\cdots,n.$$

因此,在定理 2.3 中,可以用 $\widetilde{G}_i(A,B)$ 代替 $G_i(A,B), i=1,\cdots,n$.

注 2.3. 从定理 2.3 的结论 1),可以推导出任一矩阵

$$A=(\alpha_{ij})\in \mathbf{C}^{n\times n}$$

的 Gerschgorin 圆盘 $G_i, i=1,\cdots,n$ (见第三章(2.3)). 事实上,如果 λ 是 A 的一个特征值,则对于 $t>0$, 有 $(\lambda,t)\in\lambda(A,tI)$. 据定理 2.3, 存在 i, 使得 $(\lambda,t)\in G_i(A,B)$, 即

$$\rho((\alpha_{ii},t),(\lambda,t)) \leqslant \sum_{j\neq i}|\alpha_{ij}| \Big/ \sqrt{|\alpha_{ii}|^2+t^2},$$

从而有

$$|\lambda-\alpha_{ii}| \leqslant \sqrt{1+(|\lambda|/t)^2}\cdot\sum_{j\neq i}|\alpha_{ij}|.$$

在上式中令 $t\to+\infty$, 便得到

$$\lambda \in G_i(A) = \left\{\lambda \in \mathbb{C}: |\lambda - \alpha_{ii}| \leqslant \sum_{j \neq i} |\alpha_{ij}|\right\}.$$

2.2 应用举例

设 $\{A, B\} \in \mathrm{D}_g(n)$. 据定理 1.5，存在非奇异阵

$$X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) \text{ 与 } Y = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n),$$

其中 $\boldsymbol{x}_i$ 与 $\boldsymbol{y}_i (i = 1, \cdots, n)$ 皆为单位向量，使得

$$Y^H A X = \operatorname{diag}(\alpha_1, \cdots, \alpha_n), \quad Y^H B X = \operatorname{diag}(\beta_1, \cdots, \beta_n). \tag{2.11}$$

设 εE 与 $\varepsilon F \in \mathbb{C}^{n \times n}$ 分别是 A 与 B 的微小扰动，其中 $0 < \varepsilon \ll 1$，$E = (e_{ij})$ 与 $F = (f_{ij})$ 满足

$$\sqrt{|e_{ij}|^2 + |f_{ij}|^2} \leqslant 1, \quad 1 \leqslant i, j \leqslant n. \tag{2.12}$$

令

$$Y^H E X = (\varphi_{ij}), \quad Y^H F X = (\psi_{ij}),$$

于是 (φ_{ij}) 与 (ψ_{ij}) 满足

$$\sqrt{|\varphi_{ij}|^2 + |\psi_{ij}|^2} \leqslant n, \quad 1 \leqslant i, j \leqslant n. \tag{2.13}$$

取 ε 足够小，可使 $\{A + \varepsilon E, B + \varepsilon F\}$ 是正则对.

设 (α_1, β_1) 是 $\{A, B\}$ 的 $p(1 \leqslant p \leqslant n)$ 重特征值，无妨假定它们是 $(\alpha_1, \beta_1), \cdots, (\alpha_p, \beta_p)$. 现在要问：$\{A + \varepsilon E, B + \varepsilon F\}$ 是否有 p 个靠近 (α_1, β_1) 的特征值？它们之间的距离有多大？

应用 Gerschgorin 型定理（定理 2.3），可得下述结论：

1°. $1 \leqslant p \leqslant n - 1$ 的情形. 令

$$\delta = \min_{i+1 \leqslant j \leqslant n} \rho((\alpha_1, \beta_1), (\alpha_j, \beta_j)), \tag{2.14}$$

$$\begin{aligned} \nu_1 &= \min_{1 \leqslant i \leqslant p} \sqrt{|\alpha_i|^2 + |\beta_i|^2} \\ &= \min_{1 \leqslant i \leqslant p} \sqrt{|\boldsymbol{y}_i^H A \boldsymbol{x}_i|^2 + |\boldsymbol{y}_i^H B \boldsymbol{x}_i|^2}, \end{aligned} \tag{2.15}$$

$$\nu_2 = \min_{p+1 \leqslant i \leqslant n} \sqrt{|\alpha_i|^2 + |\beta_i|^2}, \quad \nu = \min\{\nu_1, \nu_2\} \tag{2.16}$$

和

$$\chi_\varepsilon = \frac{n(p + \sqrt{\varepsilon})\varepsilon}{\nu_1 - \varepsilon}, \tag{2.17}$$

则当 ε 满足

$$\varepsilon < \frac{\nu}{2}, \quad n(p + \sqrt{\varepsilon})\varepsilon < \frac{\nu\delta}{4} \tag{2.18}$$

时，$\{A + \varepsilon E, B + \varepsilon F\}$ 恰有 p 个特征值落入非 Euclid 圆盘 $\mathscr{D}_{\chi_\varepsilon}(\alpha_1, \beta_1)$；

2°．$p = n$ 的情形．令

$$\nu = \min_{1 \leqslant i \leqslant n} \sqrt{|\alpha_i|^2 + |\beta_i|^2} \tag{2.19}$$

和

$$\chi_\varepsilon = \frac{n^2 \varepsilon}{\nu - \varepsilon}, \tag{2.20}$$

则当 $\varepsilon < \nu$ 时，$\{A + \varepsilon E, B + \varepsilon F\}$ 的 n 个特征值全部落入非 Euclid 圆盘 $\mathscr{D}_{\chi_\varepsilon}(\alpha_1, \beta_1)$．

证明如下：

1°．设 $1 \leqslant p \leqslant n - 1$．取一待定的正数 σ，用 $\varepsilon^{\frac{3}{2}}/\sigma$ 分别乘以 $\widetilde{A} = Y^H(A + \varepsilon E)X$ 与 $\widetilde{B} = Y^H(B + \varepsilon F)X$ 的第 $p + 1$ 列到第 n 列，所得到的矩阵记作 $\{\widetilde{A}, \widetilde{B}\}$，则由注 2,2 知，当 $i = 1, 2, \cdots, p$ 时，$G_i(\widetilde{A}, \widetilde{B})$ 中的点 (α, β) 必满足

$$\begin{aligned}
&\rho((\alpha_i + \varepsilon\varphi_{ii}, \beta_i + \varepsilon\phi_{ii}), (\alpha, \beta)) \\
&\leqslant \frac{\varepsilon \sum\limits_{\substack{j=1 \\ j \neq i}}^{p} \sqrt{|\varphi_{ij}|^2 + |\phi_{ij}|^2} + \dfrac{\varepsilon^{\frac{3}{2}}}{\sigma} \sum\limits_{i=p+1}^{n} \sqrt{|\varphi_{ij}|^2 + |\phi_{ij}|^2}}{\sqrt{|\alpha_i + \varepsilon\varphi_{ii}|^2 + |\beta_i + \varepsilon\phi_{ii}|^2}} \\
&\leqslant \frac{n\left[(p-1)\varepsilon + (n-p)\dfrac{\varepsilon^{\frac{3}{2}}}{\sigma}\right]}{\sqrt{|\alpha_i|^2 + |\beta_i|^2} - \varepsilon\sqrt{|\varphi_{ii}|^2 + |\phi_{ii}|^2}} \\
&\leqslant \frac{n\left[(p-1)\varepsilon + (n-p)\dfrac{\varepsilon^{\frac{3}{2}}}{\sigma}\right]}{\nu_1 - \varepsilon},
\end{aligned}$$

再利用弦度量的三角不等式，可知，$G_i(\widetilde{A}, \widetilde{B})(i = 1, \cdots, p)$ 中的点 (α, β) 满足

$$\begin{aligned}
&\rho((\alpha_i,\beta_i),(\alpha,\beta)) \\
&\quad \leqslant \rho((\alpha_i,\beta_i),(\alpha_i+\varepsilon\varphi_{ii},\beta_i+\varepsilon\psi_{ii})) \\
&\qquad +\rho((\alpha_i+\varepsilon\varphi_{ii},\beta_i+\varepsilon\psi_{ii}),(\alpha,\beta)) \\
&\quad \leqslant \frac{n\left[p\varepsilon+(n-p)\dfrac{\varepsilon^{\frac{3}{2}}}{\sigma}\right]}{\nu_1-\varepsilon} \equiv \chi_1(\varepsilon);
\end{aligned}$$

同理，当 $i=p+1,\cdots,n$ 时，$G_i(\widetilde{A},\widetilde{B})$ 中的点 (α,β) 必满足

$$\begin{aligned}
&\rho((\alpha_i+\varepsilon\varphi_{ii},\beta_i+\varepsilon\psi_{ii}),(\alpha,\beta)) \\
&\quad \leqslant \frac{\varepsilon\sum\limits_{j=1}^{p}\sqrt{|\varphi_{ij}|^2+|\psi_{ij}|^2}+\dfrac{\varepsilon^{\frac{3}{2}}}{\sigma}\sum\limits_{\substack{j=p+1\\ j\neq i}}^{n}\sqrt{|\varphi_{ij}|^2+|\psi_{ij}|^2}}{\sqrt{|\alpha_i+\varepsilon\varphi_{ii}|^2+|\beta_i+\varepsilon\psi_{ii}|^2}} \\
&\quad \leqslant \frac{n\left[p\varepsilon+(n-p-1)\dfrac{\varepsilon^{\frac{3}{2}}}{\sigma}\right]}{(\nu_2-\varepsilon)},
\end{aligned}$$

再利用三角不等式，可知 $G_i(\widetilde{A},\widetilde{B})$ $(i=p+1,\cdots,n)$ 中的点 (α,β) 满足

$$\rho((\alpha_i,\beta_i),(\alpha,\beta)) \leqslant \frac{n\left[p\varepsilon+(n-p)\dfrac{\varepsilon^{\frac{3}{2}}}{\sigma}\right]}{\nu_2-\varepsilon} \equiv \chi_2(\varepsilon).$$

今取 $\sigma=n-p$. 显然当 ε 满足条件 (2.18) 时，有

$$\chi_1(\varepsilon)+\chi_2(\varepsilon)<\frac{4n(p+\sqrt{\varepsilon})\varepsilon}{\nu}<\delta,$$

即

$$\bigcup_{i=1}^{p} G_i(\widetilde{A},\widetilde{B}) \cap \bigcup_{i=p+1}^{n} G_i(\widetilde{A},\widetilde{B}) = \emptyset.$$

因此，根据定理 2,3 的结论 2)，$\{\widetilde{A},\widetilde{B}\}$ 恰有 p 个特征值位于

$$\bigcup_{i=1}^{p} G_i(\widetilde{A},\widetilde{B})$$

之中.

再注意到

$$\bigcup_{i=1}^{p} G_i(\tilde{A},\tilde{B}) \subset \mathscr{D}_{\chi_\varepsilon}(\alpha_1,\beta_1),$$

所以，当 ε 满足条件 (2.18) 时，$\{A+\varepsilon E, B+\varepsilon F\}$ 恰有 p 个特征值落入 $\mathscr{D}_{\chi_\varepsilon}(\alpha_1,\beta_1)$.

2°. 设 $p=n$. 由注 2.2 知，对于 $\tilde{A}=Y^H(A+\varepsilon E)X$ 和 $\tilde{B}=Y^H(B+\varepsilon F)X$, $G_i(\tilde{A},\tilde{B})$ $(i=1,\cdots,n)$ 中的点 (α,β) 必满足

$$\begin{aligned}
&\rho((\alpha_i+\varepsilon\varphi_{ii},\beta_i+\varepsilon\phi_{ii}),(\alpha,\beta)) \\
&\qquad \leqslant \frac{\varepsilon\sum_{j\neq i}\sqrt{|\varphi_{ij}|^2+|\phi_{ij}|^2}}{\sqrt{|\alpha_i+\varepsilon\varphi_{ii}|^2+|\beta_i+\varepsilon\phi_{ii}|^2}} \\
&\qquad \leqslant \frac{n(n-1)\varepsilon}{\nu-\varepsilon},
\end{aligned}$$

再利用三角不等式，可知 $G_i(\tilde{A},\tilde{B})$ $(i=1,\cdots,n)$ 中的点 (α,β) 满足

$$\rho((\alpha_i,\beta_i),(\alpha,\beta)) \leqslant \frac{n^2\varepsilon}{\nu-\varepsilon} \equiv \chi_\varepsilon.$$

因此根据定理 2.3，当 $\varepsilon<\nu$ 时，有

$$\lambda(\tilde{A},\tilde{B}) \subset \bigcup_{i=1}^{n} G_i(\tilde{A},\tilde{B}) \subseteq \mathscr{D}_{\chi_\varepsilon}(\alpha_1,\beta_1),$$

即 $\{A+\varepsilon E, B+\varepsilon F\}$ 的 n 个特征值全部落入 $\mathscr{D}_{\chi_\varepsilon}(\alpha_1,\beta_1)$. □

注 2.4. 上述结论表明，如果 (α_1,β_1) 是可对角化对 $\{A,B\}$ 的 p 重特征值，则对于 (2.12) 所示的扰动 εE 和 εF，$\{A+\varepsilon E, B+\varepsilon F\}$ 必有 p 个特征值 $(\gamma_1,\delta_1),\cdots,(\gamma_p,\delta_p)$，满足

$$\rho((\alpha_1,\beta_1),(\gamma_i,\delta_i)) \lesssim \frac{np\varepsilon}{\nu_1}, \quad i=1,\cdots,p,$$

其中

$$\nu_1 = \min_{1\leqslant i\leqslant p}\sqrt{|\boldsymbol{y}_i^H A\boldsymbol{x}_i|^2+|\boldsymbol{y}_i^H B\boldsymbol{x}_i|^2},$$

$\boldsymbol{x}_1,\cdots,\boldsymbol{x}_p$ 和 $\boldsymbol{y}_1^H,\cdots,\boldsymbol{y}_p^H$ 分别是 $\{A,B\}$ 属于特征值 (α_1,β_1) 的 p 个单位右特征向量和单位左特征向量. 因此，可以把 p/ν_1 看作

$\{A,B\}$ 的 p 重特征值 (α_1,β_1) 的条件数.

注 2.5 本节取自本书作者的论文[22]. G. W. Stewart[156] 最早把 Gerschgorin 定理推广到正则对. 作者从一个一般性的命题(即定理 2.1)出发,导出了类似于 Stewart 的结果,同时消除了 Stewart 论文中的某些含混和不够严格之处.

习题

1. 设 (α,β) 是 n 阶正则对 $\{A,B\}$ 的单特征值, $\boldsymbol{x}$ 与 $\boldsymbol{x}'$ 都是 $\{A,B\}$ 属于 (α,β) 的右特征向量, 则必有 $\boldsymbol{x}'=\gamma\boldsymbol{x}$, $\gamma\in\mathbf{C}$.

2. 设 (α,β) 是 n 阶正则对 $\{A,B\}$ 的单特征值, $\boldsymbol{x}$ 与 $\boldsymbol{y}^H$ 分别是 $\{A,B\}$ 属于 (α,β) 的右特征向量与左特征向量,则 $(\boldsymbol{y}^HA\boldsymbol{x}, \boldsymbol{y}^HB\boldsymbol{x})\neq(0,0)$.

3. 设 $\alpha,\beta,\alpha_k,\beta_k\geqslant 0$, $k=1,\cdots,n$. 证明恒等式

$$\left[\sum_{k=1}^{n}(\alpha\beta_k+\beta\alpha_k)\right]^2 = (\alpha^2+\beta^2)\left[\left(\sum_{k=1}^{n}\alpha_k\right)^2+\left(\sum_{k=1}^{n}\beta_k\right)^2\right] - \left[\sum_{k=1}^{n}(\alpha\alpha_k-\beta\beta_k)\right]^2.$$

§ 3 定型对的特征值

3.1 Crawford 数 $c(A,B)$ 的性质

对于定型对 $\{A,B\}$ 来说, Crawford 数 $c(A,B)$ 是最重要的量. $c(A,B)$ 不仅出现在定义 1.5 中,而且还经常出现在定型对特征值问题的扰动界限之中.

首先证明 $c(A,B)$ 的连续性质,即

定理 3.1[159]. 设 $A,B,\widetilde{A}=A+E$ 与 $\widetilde{B}=B+F$ 均为 n 阶 Hermite 阵. 则

$$|c(\widetilde{A},\widetilde{B})-c(A,B)|\leqslant(\|E\|_2^2+\|F\|_2^2)^{\frac{1}{2}}. \tag{3.1}$$

证明:

显然只需证明

$$c(\widetilde{A},\widetilde{B}) \geqslant c(A,B) - (\|E\|_2^2 + \|F\|_2^2)^{\frac{1}{2}}. \tag{3.2}$$

利用 $c(A,B)$ 的定义(见 (1.47) 式)和不等式

$$\sqrt{(\alpha+\delta)^2+(\beta+\gamma)^2} \geqslant \sqrt{\alpha^2+\beta^2} - \sqrt{\gamma^2+\delta^2},$$
$$\alpha,\beta,\gamma,\delta \in \mathbb{R},$$

可得

$$\begin{aligned} c(\widetilde{A},\widetilde{B}) &\equiv \min_{\|x\|_2=1}\{[x^H(A+E)x]^2 + [x^H(B+F)x]^2\}^{\frac{1}{2}} \\ &\geqslant \min_{\|x\|_2=1}\{[(x^HAx)^2 + (x^HBx)^2]^{\frac{1}{2}} \\ &\quad - [(x^HEx)^2 + (x^HFx)^2]^{\frac{1}{2}}\} \\ &\geqslant \min_{\|x\|_2=1}[(x^HAx)^2 + (x^HBx)^2]^{\frac{1}{2}} \\ &\quad - [\max_{\|x\|_2=1}(x^HEx)^2 + \max_{\|x\|_2=1}(x^HFx)^2]^{\frac{1}{2}} \\ &= c(A,B) - (\|E\|_2^2 + \|F\|_2^2)^{\frac{1}{2}}. \end{aligned}$$

即不等式 (3.2) 成立。□

当 $A,B \in \mathbb{R}^{n\times n}$ 均为对称阵时,定义

$$c_r(A,B) \equiv \min_{\substack{\|x\|_2=1 \\ x\in\mathscr{R}^n}} |x^H(A+iB)x| \tag{3.3}$$

和

$$F_r(A+iB) = \{x^T(A+iB)x : x\in\mathbb{R}^n, \|x\|_2 = 1\}. \tag{3.4}$$

对于$n \geqslant 3$, Brickman[64] 证明了

$$F_r(A+iB) = F(A+iB), \tag{3.5}$$

其中 $F(A+iB)$ 表示 $A+iB$ 的值域;因此,有

$$c_r(A,B) = c(A,B). \tag{3.6}$$

可是当 $n=2$ 时,等式(3.6)一般不再成立。例如

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

有 $c(A,B) = 0 < 1 = c_r(A,B)$. 根据 $c_r(A,B)$ 与 $c(A,B)$ 的定义(见 (1.47) 与 (3.3)),一般应有

$$c(A,B) \leqslant c_r(A,B). \tag{3.7}$$

然而利用定理 1.7，可证：如果 $A, B \in \mathbb{R}^{2\times 2}$ 均为对称阵，并且 $c(A,B) > 0$，则等式(3.6)成立.

事实上，根据定理 1.7，存在 $\varphi \in [0,2\pi)$，使得(1.48)式中的 $B_\varphi > 0$，并且存在 B_φ 属于特征值 $c(A,B)$ 的一个特征向量 $\boldsymbol{x} \in \mathbf{C}^2$，$\|\boldsymbol{x}\|_2 = 1$，使得 $\boldsymbol{x}^H B_\varphi \boldsymbol{x} = c(A,B)$，同时有 $\boldsymbol{x}^H A_\varphi \boldsymbol{x} = 0$. 显然 $A_\varphi = 0$ 或者 A_φ 为不定型. 若 $A_\varphi = 0$，则必有 $\boldsymbol{x}' \in \mathbb{R}^2, \|\boldsymbol{x}'\|_2 = 1$，使得 $c(A,B) = \boldsymbol{x}'^T B_\varphi \boldsymbol{x}' \geqslant c_r(A,B)$，因而 (3.6) 式成立. 以下假设 A_φ 为不定型. 如果 $c(A,B)$ 是 B_φ 的单特征值，则 $\boldsymbol{x}$ 必可表示成 $\boldsymbol{x} = p\boldsymbol{x}'$，其中 $\boldsymbol{x}' \in \mathbb{R}^2$，$p \in \mathbf{C}$；$\boldsymbol{x}'$ 满足 $\boldsymbol{x}'^T A_\varphi \boldsymbol{x}' = 0$ 和 $c(A,B) = \boldsymbol{x}'^T B_\varphi \boldsymbol{x}' \geqslant c_r(A,B)$，所以 (3.6) 式成立. 如果 $c(A, B)$ 是 B_φ 的重特征值，则任一非零向量都是 B_φ 的特征向量；因为实对称阵 A_φ 为不定型，所以必有向量 $\boldsymbol{x}' \in \mathbb{R}^2$，$\|\boldsymbol{x}'\|_2 = 1$，使得 $\boldsymbol{x}'^T A_\varphi \boldsymbol{x}' = 0$，同时有 $c(A, B) = \boldsymbol{x}'^T B_\varphi \boldsymbol{x}' \geqslant c_r(A, B)$，即等式 (3.6) 亦成立.

因此，这就证明了

定理 3.2. 设 n 为任一自然数，$A, B \in \mathbb{R}^{n\times n}$，并且 $\{A,B\} \in \mathbb{D}(n)$. 则等式 (3.6) 成立.

3.2 D(n) 上的一种投影度量

定义 3.1. 设 $\{A,B\},\{C,D\} \in \mathbb{D}(n)$，如果存在 $r \in \mathbb{R}$ 使得

$$\left.\begin{array}{r}(C,D) = (rA, rB),\ \text{或}\ (A,C) = (rB, rD),\\ \text{或}\ (B,D) = (rA, rC),\end{array}\right\} \tag{3.8}$$

则称 $\{A,B\}$ 与 $\{C,D\}$ 等价，记作 $\{A,B\} \sim \{C,D\}$.

按照定义 3.1，可以将 $\mathbb{D}(n)$ 的元素划分成不同的等价类；如果把每一个等价类看作一点，则由所有这样的点构成一个投影空间，记作 $\mathbf{G}(n)$.

设 $\{A,B\},\{C,D\} \in \mathbf{D}(n)$. 定义

$$s(\{A,B\},\{C,D\}) = \max_{\|\boldsymbol{x}\|_2=1} \rho((\boldsymbol{x}^H A\boldsymbol{x}, \boldsymbol{x}^H B\boldsymbol{x}), (\boldsymbol{x}^H C\boldsymbol{x}, \boldsymbol{x}^H D\boldsymbol{x})). \tag{3.9}$$

定理 3.3[11]**.** (3.9) 式所定义的 $s(\{A,B\},\{C,D\})$ 是投影空

间 $\mathbf{G}(n)$ 上的度量.

证明:

要证的是,如果 $\{A,B\},\{C,D\}$ 和 $\{E,F\}\in\mathbf{D}(n)$, 则

(i) $s(\{A,B\},\{C,D\})\geqslant 0$,并且当且仅当 $\{C,D\}\sim\{A,B\}$ 时,有 $s(\{A,B\},\{C,D\})=0$;

(ii) $s(\{A,B\},\{C,D\})=s(\{C,D\},\{A,B\})$;

(iii) $s(\{A,B\},\{C,D\})\leqslant s(\{A,B\},\{E,F\})$
$+s(\{E,F\},\{C,D\})$.

利用弦度量 $\rho(\cdot,\cdot)$ 的度量性质以及定义 (3.9),容易推导出上述性质 (ii)、(iii) 和 (i) 中的第一个断言,并且容易从定义 3.1 看出,当 $\{C,D\}\sim\{A,B\}$时, 有 $s(\{A,B\},\{C,D\})=0$. 以下证明: 对于 $\{A,B\},\{C,D\}\in\mathbf{D}(n)$, 如果

$s(\{A,B\},\{C,D\})=0$, 则 $\{C,D\}\sim\{A,B\}$.

$s(\{A,B\},\{C,D\})=0$ 表示

$$\boldsymbol{x}^H A\boldsymbol{x}\boldsymbol{x}^H D\boldsymbol{x}=\boldsymbol{x}^H B\boldsymbol{x}\boldsymbol{x}^H C\boldsymbol{x},\quad \forall\boldsymbol{x}\in\mathbf{C}^n.$$

如果存在 $r\in\mathbf{R}, A=rB$, 则从 $\{A,B\}\in\mathbf{D}(n)$ 可知 $\boldsymbol{x}^H B\boldsymbol{x}\neq 0$, $\forall\boldsymbol{x}\in\mathbf{C}^n$, $\boldsymbol{x}\neq 0$. 于是有

$$\boldsymbol{x}^H C\boldsymbol{x}=r\boldsymbol{x}^H D\boldsymbol{x},\quad \forall\boldsymbol{x}\in\mathbf{C}^n.$$

因此 $C=rD$, 即 $\{C,D\}\sim\{A,B\}$. 类似地, 如果存在 $r\in\mathbf{R}$, $B=rA$, 则 $D=rC$, 这时也得到 $\{C,D\}\sim\{A,B\}$. 所以在下面的证明中,将假定对一切 $r\in\mathbf{R}, A\neq rB$, 并且 $B\neq rA$. 证明分下列 4 步:

1) 根据定理 1.7, 存在 $\varphi\in[0,2\pi)$, 使得

$$B_\varphi=\sin\varphi A+\cos\varphi B>0.$$

令

$$A_\varphi=\cos\varphi A-\sin\varphi B,\quad C_\varphi=\cos\varphi C-\sin\varphi D$$

和

$$D_\varphi=\sin\varphi C+\cos\varphi D,$$

则有

$$\boldsymbol{x}^H A_\varphi\boldsymbol{x}\boldsymbol{x}^H D_\varphi\boldsymbol{x}=\boldsymbol{x}^H B_\varphi\boldsymbol{x}\boldsymbol{x}^H C_\varphi\boldsymbol{x},\quad \forall\boldsymbol{x}\in\mathbf{C}^n.$$

2) 分解 $B_\varphi = H^2$,其中 $H > 0$. 令 $A_\varphi = H\hat{A}H, C_\varphi = H\hat{C}H$, $D_\varphi = H\hat{D}H$ 和 $\boldsymbol{y} = H\boldsymbol{x}$, 则有

$$\boldsymbol{y}^H \hat{A} \boldsymbol{y}\boldsymbol{y}^H \hat{D} \boldsymbol{y} = \boldsymbol{y}^H \boldsymbol{y}\boldsymbol{y}^H \hat{C} \boldsymbol{y}, \quad \forall \boldsymbol{y} \in \mathbb{C}^n.$$

3) 分解 $\hat{A} = V^H \Lambda V$, 其中 V 为酉阵,并且

$$\Lambda = \mathrm{diag}(\alpha_1 I^{(n_1)}, \cdots, \alpha_m I^{(n_m)}), \quad \alpha_k \neq \alpha_l \ (k \neq l).$$

然后令 $\boldsymbol{z} = V\boldsymbol{y}, \hat{C} = V^H C' V$ 和 $\hat{D} = V^H D' V$, 则有

$$\boldsymbol{z}^H \Lambda \boldsymbol{z}\boldsymbol{z}^H D' \boldsymbol{z} = \boldsymbol{z}^H \boldsymbol{z}\boldsymbol{z}^H C' \boldsymbol{z}, \quad \forall \boldsymbol{z} \in \mathbb{C}^n. \tag{3.10}$$

4) 记

$$C' = \begin{pmatrix} C_{11} \cdots C_{1m} \\ \cdots \\ C_{m1} \cdots C_{mm} \end{pmatrix}, \quad D' = \begin{pmatrix} D_{11} \cdots D_{1m} \\ \cdots \\ D_{m1} \cdots D_{mm} \end{pmatrix},$$

其中 C_{kl} 和 $D_{kl} \in \mathbb{C}^{n_k \times n_l}$, $k, l = 1, \cdots, m$, $\sum_{k=1}^{m} n_k = n$.

令

$$\boldsymbol{z} = (\boldsymbol{z}_1^H, \cdots, \boldsymbol{z}_m^H)^H \in \mathbb{C}^n,$$

其中 $\boldsymbol{z}_k \in \mathbb{C}^{n_k}$, $1 \leqslant k \leqslant m$. 对任一自然数 $k(1 \leqslant k \leqslant m)$ 和 $t \in \mathbb{R}$, 取 $\boldsymbol{z} = (\boldsymbol{z}_1^H, 0, \cdots, 0, t\boldsymbol{z}_k^H, 0, \cdots, 0)^H \in \mathbb{C}^n$, 代入 (3.10), 得到

$$\begin{aligned}(\alpha_1 \boldsymbol{z}_1^H \boldsymbol{z}_1 + \alpha_k \boldsymbol{z}_k^H \boldsymbol{z}_k t^2)[\boldsymbol{z}_1^H D_{11} \boldsymbol{z}_1 + (\boldsymbol{z}_1^H D_{1k} \boldsymbol{z}_k + \boldsymbol{z}_k^H D_{k1} \boldsymbol{z}_1)t + \boldsymbol{z}_k^H D_{kk} \boldsymbol{z}_k t^2] \\ = (\boldsymbol{z}_1^H \boldsymbol{z}_1 + \boldsymbol{z}_k^H \boldsymbol{z}_k t^2)[\boldsymbol{z}_1^H C_{11} \boldsymbol{z}_1 + (\boldsymbol{z}_1^H C_{1k} \boldsymbol{z}_k + \boldsymbol{z}_k^H C_{k1} \boldsymbol{z}_1)t \\ + \boldsymbol{z}_k^H C_{kk} \boldsymbol{z}_k t^2], \quad \forall \boldsymbol{z}_1 \in \mathbb{C}^{n_1}, \boldsymbol{z}_k \in \mathbb{C}^{n_k}, t \in \mathbb{R}.\end{aligned}$$

比较上式两端 t 的同次幂的系数,可知

$$\alpha_j \boldsymbol{z}_j^H D_{jj} \boldsymbol{z}_j = \boldsymbol{z}_j^H C_{jj} \boldsymbol{z}_j, \quad \forall \boldsymbol{z}_j \in \mathbb{C}^{n_j}, \quad j = 1, k; \tag{3.11}$$

$$\boldsymbol{z}_1^H \boldsymbol{z}_1 \boldsymbol{z}_k^H (\alpha_1 D_{kk} - C_{kk}) \boldsymbol{z}_k = \boldsymbol{z}_k^H \boldsymbol{z}_k \boldsymbol{z}_1^H (C_{11} - \alpha_k D_{11}) \boldsymbol{z}_1, \quad \forall \boldsymbol{z}_j \in \mathbb{C}^{n_j}, \quad j = 1, k; \tag{3.12}$$

$$\alpha_l (\boldsymbol{z}_1^H D_{1k} \boldsymbol{z}_k + \boldsymbol{z}_k^H D_{k1} \boldsymbol{z}_1) = \boldsymbol{z}_1^H C_{1k} \boldsymbol{z}_k + \boldsymbol{z}_k^H D_{k1} \boldsymbol{z}_1, \quad l = 1, k, \ \forall \boldsymbol{z}_j \in \mathbb{C}^{n_j}, j = 1, k. \tag{3.13}$$

从 (3.11) 得到

$$C_{11} = \alpha_1 D_{11}, \quad C_{kk} = \alpha_k D_{kk}.$$

代入 (3.12), 得到

$$\frac{\boldsymbol{z}_k^H D_{kk} \boldsymbol{z}_k}{\boldsymbol{z}_k^H \boldsymbol{z}_k} = \frac{\boldsymbol{z}_1^H D_{11} \boldsymbol{z}_1}{\boldsymbol{z}_1^H \boldsymbol{z}_1} = r \in \mathbb{R}, \ \forall \boldsymbol{z}_j \in \mathbb{C}^{n_j}, \ j = 1, k,$$

从而

$$D_{11}=D_{kk}=rI.$$

由 (3.13) 可知

$$z_1^H D_{1k} z_k + z_k^H D_{k1} z_1 = 0, \quad \forall z_j \in \mathbb{C}^{n_j},\ j=1,k. \tag{3.14}$$

在 (3.14) 式左端分别取 $z_k = D_{1k}^H z_1$ 和 $z_1 = D_{k1}^H z_k$，得出

$$D_{1k}=0, \quad D_{k1}=0.$$

因此，可导出

$$D'=rI, \quad C'=rA, \quad r\in\mathbb{R}.$$

综合上述1)—4)，可知存在 $r\in\mathbb{R}$，$r\neq 0$，使

$$(C,D)=(rA,rB),$$

即

$$\{C,D\}\sim\{A,B\}. \qquad \square$$

3.3 Weyl-Лидский 型定理

设 $\{A,B\}\in \mathrm{D}(n)$. 如果 Hermite 阵 E 与 $F\in\mathbb{C}^{n\times n}$ 满足 $\sqrt{\|E\|_2^2+\|F\|_2^2}<c(A,B)$，令 $\widetilde{A}=A+E$ 和 $\widetilde{B}=B+F$，则由定理 3.1 可知 $\{\widetilde{A},\widetilde{B}\}\in\mathrm{D}(n)$.

根据定理 1.7 的证明(见图 4-1)，$A+iB$ 的值域 $F(A+iB)$ 位于 C 内一个不包含原点的半平面 $\mathscr{H}$ 之中；同理，$\widetilde{A}+i\widetilde{B}$ 的值域 $F(\widetilde{A}+i\widetilde{B})$ 位于 C 内一个不包含原点的半平面 $\widetilde{\mathscr{H}}$ 之中. 于是，在 $\mathbb{C}\backslash(\mathscr{H}\cup\widetilde{\mathscr{H}})$ 内必含有一条从原点发出的射线 R（如图 4-2 所示）. 对 C 中的每个非零点 $\xi+i\eta$，定义一个角 $\theta(\xi,\eta)$，它是射线 R 和射线 $\{\tau(\xi+i\eta):\ \tau\geqslant 0\}$ 之间按顺时针方向的夹角. 显然，$\theta(\xi,\eta)$ 是 $\mathscr{H}\cup\widetilde{\mathscr{H}}$ 上的连续函数.

设 $\{A,B\}\in\mathrm{D}(n)$，$\lambda(A,B)=\{(\alpha_k,\beta_k)\}_{k=1}^n$. 根据定义 1.2，存在 $x_k\in\mathbb{C}^n$，$x_k\neq 0$，使得

$$\beta_k A x_k = \alpha_k B x_k, \quad k=1,\cdots,n.$$

现定义与特征值 (α_k,β_k) 相联系的角

$$\theta_k \equiv \theta(x_k^H A x_k, x_k^H B x_k), \quad k=1,\cdots,n. \tag{3.15}$$

$\theta_1,\cdots,\theta_n$ 叫做 $\{A,B\}$ 的特征角.

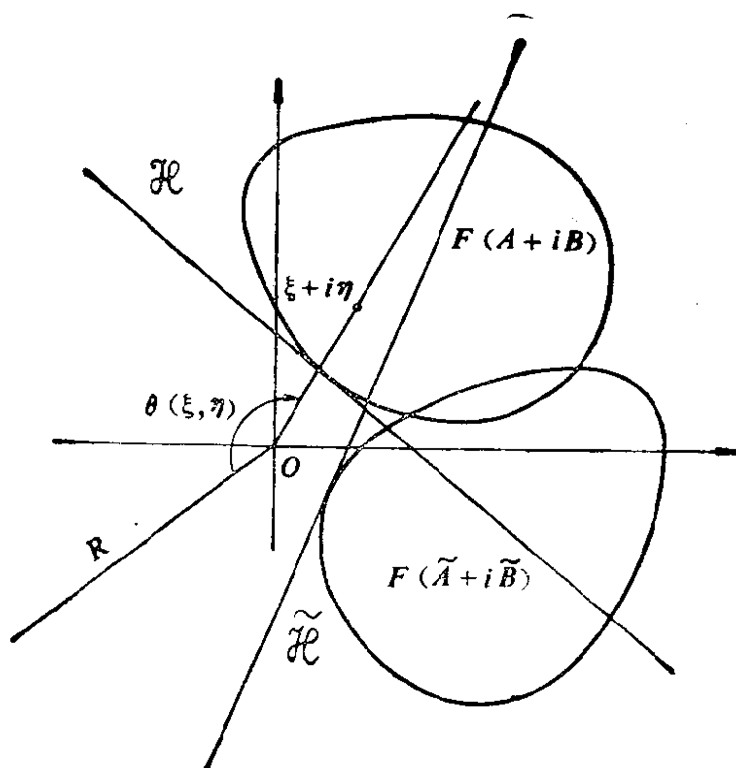


图 4-2

注意到，由 $\beta_k \boldsymbol{x}_k^H A \boldsymbol{x}_k = \alpha_k \boldsymbol{x}_k^H B \boldsymbol{x}_k$ 可知，存在非零实数 σ，使得

$$(\alpha_k, \beta_k) = \sigma(\boldsymbol{x}_k^H A \boldsymbol{x}_k, \boldsymbol{x}_k^H B \boldsymbol{x}_k);$$

所以

$$\theta_k = \theta(\alpha_k, \beta_k), \quad k = 1, \cdots, n. \tag{3.16}$$

这些特征角 θ_k，具有类似于 Hermite 阵特征值的 minimax 性质，即有

定理3.4[159]. 设 $\{A, B\} \in \mathbb{D}(n)$，$\{\theta_k\}$ 是 $\{A, B\}$ 的特征角，并且排列为 $\theta_1 \leqslant \cdots \leqslant \theta_n$，则有

$$\theta_k = \min_{\substack{\mathscr{X} \subset \mathbb{C}^n \\ \dim(\mathscr{X}) = k}} \max_{\substack{\boldsymbol{x} \in \mathscr{X} \\ \boldsymbol{x} \neq 0}} \theta(\boldsymbol{x}^H A \boldsymbol{x}, \boldsymbol{x}^H B \boldsymbol{x}) \tag{3.17}$$

和

$$\theta_k = \max_{\substack{\mathscr{X} \subset \mathbb{C}^n \\ \dim(\mathscr{X}) = n-k+1}} \min_{\substack{\boldsymbol{x} \in \mathscr{X} \\ \boldsymbol{x} \neq 0}} \theta(\boldsymbol{x}^H A \boldsymbol{x}, \boldsymbol{x}^H B \boldsymbol{x}), \tag{3.18}$$

$$k = 1, \cdots, n.$$

证明:

首先取 $\varphi \in [0, 2\pi)$，使经变换 (1.48) 后，所得到的 $B_\varphi > 0$.

与此同时，将定义 θ 的射线 R 按反时针方向旋转 φ 角，并利用这条新的射线定义新的函数 θ_φ，则有

$$\theta_\varphi(x^H A_\varphi x, x^H B_\varphi x) = \theta(x^H A x, x^H B x).$$

因此，不失一般性，可以假定 $B > 0$.

容易证明(本节习题 1)，如果 $A, B \in \mathbb{C}^{n\times n}$ 均为 Hermite 阵，$B > 0$，并且 $\lambda_1 \leqslant \cdots \leqslant \lambda_n$ 是 $\{A, B\}$ 的特征值，则

$$\lambda_k = \min_{\substack{\mathscr{X}\subset\mathbb{C}^n \\ \dim(\mathscr{X})=k}} \max_{\substack{x\in\mathscr{X} \\ x\neq 0}} \frac{x^H A x}{x^H B x}, \quad k = 1, \cdots, n.$$

因为映射 $\lambda \to \theta(\lambda, 1)$ 在上半平面是递增的，所以

$$\theta_k = \theta(\lambda_k, 1) = \min_{\substack{\mathscr{X}\subset\mathbb{C}^n \\ \dim(\mathscr{X})=k}} \max_{\substack{x\in\mathscr{X} \\ x\neq 0}} \theta\left(\frac{x^H A x}{x^H B x}, 1\right)$$

$$= \min_{\substack{\mathscr{X}\subset\mathbb{C}^n \\ \dim(\mathscr{X})=k}} \max_{\substack{x\in\mathscr{X} \\ x\neq 0}} \theta(x^H A x, x^H B x).$$

同理从

$$\lambda_k = \max_{\substack{\mathscr{X}\subset\mathbb{C}^n \\ \dim(\mathscr{X})=n-k+1}} \min_{\substack{x\in\mathscr{X} \\ x\neq 0}} \frac{x^H A x}{x^H B x}$$

可得等式 (3.18). □

引理 3.1. 设 $\alpha, \beta, \tilde{\alpha}, \tilde{\beta} \in \mathbb{R}$，$(\alpha, \beta)$ 与 $(\tilde{\alpha}, \tilde{\beta}) \in \mathbb{G}_{1,2}$. 令 $z = \alpha + i\beta$，$\tilde{z} = \tilde{\alpha} + i\tilde{\beta}$. 如果从原点发出，分别经过点 z 与点 $\tilde{z}$ 的两条射线之间的夹角 $\theta \leqslant \pi$，则

$$\sin\theta = \rho((\alpha, \beta), (\tilde{\alpha}, \tilde{\beta})). \tag{3.19}$$

证明:

令 $v = (\alpha, \beta)^T \in \mathbb{R}^2$，$\tilde{v} = (\tilde{\alpha}, \tilde{\beta})^T \in \mathbb{R}^2$. 显然有

$$\theta = \arccos \frac{v^T \tilde{v}}{\|v\|_2 \|\tilde{v}\|_2}, \quad 0 \leqslant \theta \leqslant \pi.$$

于是

$$\sin\theta = \sqrt{1 - \cos^2\theta} = \sqrt{1 - \frac{(v^T\tilde{v})^2}{\|v\|_2^2 \|\tilde{v}\|_2^2}}.$$

将 v 与 $\tilde{v}$ 的坐标分量代入，即可得出等式 (3.19). □

下述结果是 Weyl-Лидский 定理的推广.

定理 3.5[11]. 设 $\{A, B\} \in \mathbb{D}(n)$，$\tilde{A} = A + E$ 与 $\tilde{B} = B + F$

$\in \mathbf{C}^{n\times n}$ 均为 Hermite 阵. 那末

1) 如果

$$\delta(\{A,B\},\{\widetilde{A},\widetilde{B}\}) \equiv \max_{\|\boldsymbol{x}\|_2=1}\sqrt{\frac{(\boldsymbol{x}^H E\boldsymbol{x})^2+(\boldsymbol{x}^H F\boldsymbol{x})^2}{(\boldsymbol{x}^H A\boldsymbol{x})^2+(\boldsymbol{x}^H B\boldsymbol{x})^2}} < 1, \quad (3.20)$$

则 $\{\widetilde{A},\widetilde{B}\}\in \mathrm{D}(n)$;

2) 对于 $\mathrm{D}(n)$ 中上述的矩阵对 $\{A, B\}$ 与 $\{\widetilde{A}, \widetilde{B}\}$, 如果 $\lambda(A,B)=\{(\alpha_i,\beta_i)\}$ 与 $\lambda(\widetilde{A},\widetilde{B})=\{(\tilde{\alpha}_i,\tilde{\beta}_i)\}$, 并且这些特征值分别排列得使相应的特征角 $\{\theta_k\}$ 与 $\{\tilde{\theta}_k\}$ 满足 $\theta_1\leqslant\cdots\leqslant\theta_n$ 与 $\tilde{\theta}_1\leqslant\cdots\leqslant\tilde{\theta}_n$, 则

$$\rho((\alpha_k,\beta_k)(\tilde{\alpha}_k,\tilde{\beta}_k))\leqslant s(\{A,B\},\{\widetilde{A},\widetilde{B}\}),$$
$$k=1,\cdots,n. \quad (3.21)$$

其中 $s(\{A,B\},\{\widetilde{A},\widetilde{B}\})$ 如(3.9)式所示.

证明:

1) 利用 Crawford 数的定义和 $\{A,B\}\in\mathrm{D}(n)$ 以及条件(3.20), 可知

$$\begin{aligned}
c(\widetilde{A},\widetilde{B}) &\equiv \min_{\|\boldsymbol{x}\|_2=1}\{[\boldsymbol{x}^H(A+E)\boldsymbol{x}]^2+[\boldsymbol{x}^H(B+F)\boldsymbol{x}]^2\}^{\frac{1}{2}}\\
&\geqslant \min_{\|\boldsymbol{x}\|_2=1}\{\sqrt{(\boldsymbol{x}^H A\boldsymbol{x})^2+(\boldsymbol{x}^H B\boldsymbol{x})^2}\\
&\quad -\sqrt{(\boldsymbol{x}^H E\boldsymbol{x})^2+(\boldsymbol{x}^H F\boldsymbol{x})^2}\}\\
&\geqslant \min_{\|\boldsymbol{x}\|_2=1}\sqrt{(\boldsymbol{x}^H A\boldsymbol{x})^2+(\boldsymbol{x}^H B\boldsymbol{x})^2}\\
&\quad \times\left\{1-\max_{\|\boldsymbol{x}\|_2=1}\sqrt{\frac{(\boldsymbol{x}^H E\boldsymbol{x})^2+(\boldsymbol{x}^H F\boldsymbol{x})^2}{(\boldsymbol{x}^H A\boldsymbol{x})^2+(\boldsymbol{x}^H B\boldsymbol{x})^2}}\right\}\\
&= c(A,B)(1-\delta(\{A,B\},\{\widetilde{A},\widetilde{B}\}))>0.
\end{aligned}$$

因此 $\{\widetilde{A},\widetilde{B}\}\in\mathrm{D}(n)$.

2) 对任一固定的自然数 $k(1\leqslant k\leqslant n)$, 假定 $\tilde{\theta}_k\geqslant\theta_k$. 设 $\mathscr{X}_k$ 是(3.17)式中的极小化子空间,则有

$$\tilde{\theta}_k\leqslant\max_{\substack{\boldsymbol{x}\in\mathscr{X}_k\\ \boldsymbol{x}\neq 0}}\theta(\boldsymbol{x}^H\widetilde{A}\boldsymbol{x},\boldsymbol{x}^H\widetilde{B}\boldsymbol{x}). \quad (3.22)$$

设(3.22)式右端的 $\theta(\boldsymbol{x}^H\widetilde{A}\boldsymbol{x},\boldsymbol{x}^H\widetilde{B}\boldsymbol{x})$ 在 $\mathscr{X}_k$ 中的单位向量 $\boldsymbol{x}$ 上达到最大值. 于是

$$\tilde{\theta}_k \leqslant \theta(\boldsymbol{x}^H\widetilde{A}\boldsymbol{x},\boldsymbol{x}^H\widetilde{B}\boldsymbol{x}),\theta(\boldsymbol{x}^HA\boldsymbol{x},\boldsymbol{x}^HB\boldsymbol{x}) \leqslant \theta_k, \qquad (3.23)$$

并且根据题设条件(3.20)，对于点 $Z:(\boldsymbol{x}^HA\boldsymbol{x},\boldsymbol{x}^HB\boldsymbol{x})$ 和点 $\widetilde{Z}:(\boldsymbol{x}^H\widetilde{A}\boldsymbol{x},\boldsymbol{x}^H\widetilde{B}\boldsymbol{x})$（见图 4-3)，有$|Z\widetilde{Z}| < |OZ|$. 用 P 表示点 Z 到直线 $O\widetilde{Z}$ 上的垂线的垂足，则初等几何给出 $0 \leqslant \psi < \frac{\pi}{2}$ 和

$$\begin{aligned}\tilde{\theta}_k - \theta_k \leqslant \psi &= \sin^{-1}\sqrt{1-\left(\frac{|OP|}{|OZ|}\right)^2}\\ &= \sin^{-1}\sqrt{1-\frac{(\boldsymbol{x}^HA\boldsymbol{x}\boldsymbol{x}^H\widetilde{A}\boldsymbol{x}+\boldsymbol{x}^HB\boldsymbol{x}\boldsymbol{x}^H\widetilde{B}\boldsymbol{x})^2}{[(\boldsymbol{x}^HA\boldsymbol{x})^2+(\boldsymbol{x}^HB\boldsymbol{x})^2][(\boldsymbol{x}^H\widetilde{A}\boldsymbol{x})^2+(\boldsymbol{x}^H\widetilde{B}\boldsymbol{x})^2]}}\\ &= \sin^{-1}\rho((\boldsymbol{x}^HA\boldsymbol{x},\boldsymbol{x}^HB\boldsymbol{x}),(\boldsymbol{x}^H\widetilde{A}\boldsymbol{x},\boldsymbol{x}^H\widetilde{B}\boldsymbol{x})). \qquad (3.24)\end{aligned}$$

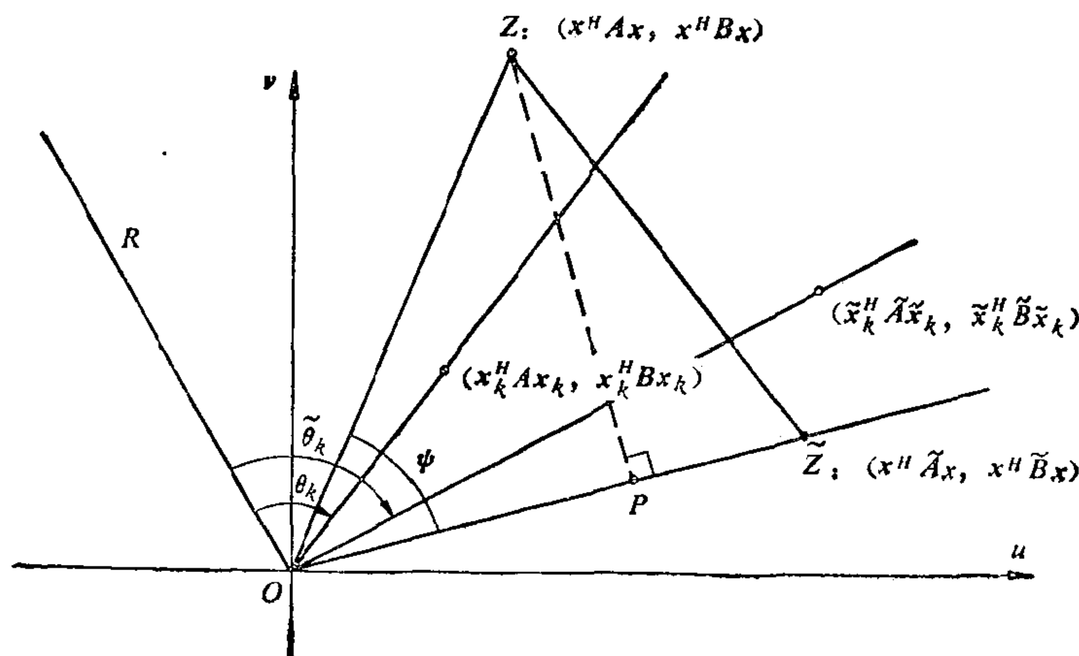


图 4-3

当 $\tilde{\theta}_k \leqslant \theta_k$ 时，可以利用(3,18)式，通过类似于上面的论证，得知存在单位向量 $\boldsymbol{x}' \in \mathbf{C}^n$，使得

$$\theta_k - \tilde{\theta}_k \leqslant \sin^{-1}\rho((\boldsymbol{x}'^HA\boldsymbol{x}',\boldsymbol{x}'^HB\boldsymbol{x}'),(\boldsymbol{x}'^H\widetilde{A}\boldsymbol{x}',\boldsymbol{x}'^H\widetilde{B}\boldsymbol{x}'))。 \qquad (3.25)$$

于是，由不等式(3.24)和(3.25)得到

$$|\theta_k - \tilde{\theta}_k| \leqslant \sin^{-1}s(\{A,B\},\{\widetilde{A},\widetilde{B}\}),$$

或者等价的$\left(因为\ 0 \leqslant |\theta_k - \tilde{\theta}_k| < \frac{\pi}{2}\right)$

$$\sin|\theta_k - \tilde{\theta}_k| \leqslant s(\{A,B\},\{\widetilde{A},\widetilde{B}\}), \quad k=1,\cdots,n.$$

再利用引理 3.1，由上列不等式立即导出不等式(3.21). □

注意到

$$s(\{A,B\},\{\tilde{A},\tilde{B}\}) \leqslant \delta(\{A,B\},\{\tilde{A},\tilde{B}\}) \leqslant \frac{\sqrt{\|E\|_2^2+\|F\|_2^2}}{c(A,B)}, \quad (3.26)$$

所以从定理 3.5 可得下述推论.

推论 3.1[159]. 设 $\{A,B\}\in\mathbb{D}(n)$，$\tilde{A}=A+E$ 与 $\tilde{B}=B+F\in\mathbf{C}^{n\times n}$ 均为 Hermite 阵. 那末，

1) 如果

$$\frac{\sqrt{\|E\|_2^2+\|F\|_2^2}}{c(A,B)}<1, \quad (3.27)$$

则 $\{\tilde{A},\tilde{B}\}\in\mathbb{D}(n)$；

2) 对于 $\mathbb{D}(n)$ 中上述的矩阵对 $\{A, B\}$ 与 $\{\tilde{A}, \tilde{B}\}$，如果 $\lambda(A,B)=\{(\alpha_k,\beta_k)\}$ 与 $\lambda(\tilde{A},\tilde{B})=\{(\tilde{\alpha}_k,\tilde{\beta}_k)\}$，并且这些特征值排列得使相应的特征角 $\{\theta_k\}$ 与 $\{\tilde{\theta}_k\}$ 满足 $\theta_1\leqslant\cdots\leqslant\theta_n$ 与 $\tilde{\theta}_1\leqslant\cdots\leqslant\tilde{\theta}_n$，则

$$\rho((\alpha_k,\beta_k),(\tilde{\alpha}_k,\tilde{\beta}_k))\leqslant\frac{\sqrt{\|E\|_2^2+\|F\|_2^2}}{c(A,B)}, k=1,\cdots,n. \quad (3.28)$$

注 3.1 由不等式(3.26)可知，推论 3.1 的条件(3.27)比定理 3.5的条件(3.20)强，而结论(3.28)比(3.21)弱. 现举一例，说明这个差异. 设 $\{A,B\}\in\mathbb{D}(n)$. 考虑

$$\{\tilde{A},\tilde{B}\}=\{(1+r)A,(1+r)B\},$$

其中 $r>0$ 满足

$$\frac{r\sqrt{\|A\|_2^2+\|B\|_2^2}}{c(A,B)}<1, \quad \frac{r\sqrt{\|A\|_2^2+\|B\|_2^2}}{c(A,B)}\approx 1.$$

显然 $\{\tilde{A},\tilde{B}\}\in\mathbb{D}(n)$，并且 $\lambda(\tilde{A},\tilde{B})=\lambda(A,B)$. 有

$$\rho((\alpha_k,\beta_k),(\tilde{\alpha}_k,\tilde{\beta}_k))=s(\{A,B\},\{\tilde{A},\tilde{B}\})=0, \ 1\leqslant k\leqslant n,$$

而

$$\frac{\sqrt{\|E\|_2^2+\|F\|_2^2}}{c(A,B)}=\frac{r\sqrt{\|A\|_2^2+\|B\|_2^2}}{c(A,B)}\approx 1.$$

注 3.2　由定理 3.5 可以导出 Weyl-Лидский 定理（第三章定理 3.6）。事实上，如果 A 与 $\widetilde{A}=A+E\in\mathbf{C}^{n\times n}$ 均为 Hermite 阵，$\lambda(A)=\{\alpha_k\}$，$\lambda(\widetilde{A})=\{\tilde{\alpha}_k\}$，并且满足 $\alpha_1\leqslant\cdots\leqslant\alpha_n$ 和 $\tilde{\alpha}_1\leqslant\cdots\leqslant\tilde{\alpha}_n$。则当 $t>0$ 时，$\{A,tI\}$ 与 $\{\widetilde{A},tI\}\in\mathrm{D}(n)$，

$$\lambda(A,tI)=\{(\alpha_k,t)\},\ \lambda(\widetilde{A},tI)=\{(\tilde{\alpha}_k,t)\},$$

并且相应的特征角 $\{\theta_k\}$ 与 $\{\tilde{\theta}_k\}$ 满足 $\theta_1\leqslant\cdots\leqslant\theta_n$ 与 $\tilde{\theta}_1\leqslant\cdots\leqslant\tilde{\theta}_n$。易知，当 $t>0$ 足够大时，

$$\delta(\{A,tI\},\{\widetilde{A},tI\})\equiv\max_{\|x\|_2=1}\frac{x^HEx}{\sqrt{(x^HAx)^2+t^2}}<1,$$

即条件 (3.20) 被满足。因此对于足够大的 $t>0$，有

$$\rho((\alpha_k,t),(\tilde{\alpha}_k,t))\leqslant\max_{\|x\|_2=1}\rho((x^HAx,t),(x^H\widetilde{A}x,t)),$$

即

$$\frac{|\alpha_k-\tilde{\alpha}_k|}{\sqrt{\left(1+\left(\frac{\alpha_k}{t}\right)^2\right)\left(1+\left(\frac{\tilde{\alpha}_k}{t}\right)^2\right)}}$$

$$\leqslant\max_{\|x\|_2=1}\frac{|x^H(A-\widetilde{A})x|}{\sqrt{\left(1+\left(\frac{x^HAx}{t}\right)^2\right)\left(1+\left(\frac{x^H\widetilde{A}x}{t}\right)^2\right)}}.$$

令 $t\to+\infty$，便得到

$$|\alpha_k-\tilde{\alpha}_k|\leqslant\max_{\|x\|_2=1}|x^H(A-\widetilde{A})x|=\|A-\widetilde{A}\|_2,$$

$$k=1,\cdots,n.$$

这正是 Weyl-Лидский 定理的结论。

3.4　关于广义奇异值的扰动

这里只是简略地综述关于广义奇异值的扰动性质。有关广义奇异值问题的研究，读者可参阅 Van Loan, Paige, Saunders, Stewart 以及作者的论文 [10]，[15]、[17]、[134]、[135]、[161]和[175]。

设 $A\in\mathbf{C}^{m\times n}$，$B\in\mathbf{C}^{p\times n}$。如果 $\mathrm{rank}\begin{pmatrix}A\\B\end{pmatrix}=n$，则称 $\{A,B\}$

是一个 $(m, p; n)$ 矩阵对，所有 $(m, p; n)$ 矩阵对的集合，记作 $\mathbf{P}(m,p;n)$. 容易看出，如果 $\{A,B\}\in\mathbf{P}(m,p;n)$，则 $\{A^HA, B^HB\}\in\mathbf{D}(n)$. 此外，如果 $\{H,K\}\in\mathbf{D}(n)$，并且 $H\geqslant 0$, $K\geqslant 0$，则据定理 1,8，总可以假定 $\{H,K\}$ 的 n 个特征值 (λ_k, μ_k) 适合 $\lambda_k\geqslant 0$, $\mu_k\geqslant 0$, $k=1,\cdots,n$. 因此建议下述定义.

定义 3.2. 设 $\{A, B\}\in\mathbf{P}(m, p; n)$. 如果 $(\lambda, \mu)\in\lambda(A^HA, B^HB)$，并且 $\lambda\geqslant 0$, $\mu\geqslant 0$，则 $(\alpha,\beta)=(\lambda^{\frac{1}{2}}, \mu^{\frac{1}{2}})$ 叫做 $\{A,B\}$ 的广义奇异值.

显然，任一 $\{A,B\}\in\mathbf{P}(m,p;n)$ 有 n 个广义奇异值，它们分布在 $\mathbf{G}_{1,2}$ 上. $\{A,B\}$ 的 n 个广义奇异值的全体，记作 $\sigma(A,B)$.

利用定理 1.8，容易证明下述关于 $(m,p;n)$ 矩阵对 $\{A,B\}$ 的广义奇异值分解定理（本节习题 5）.

定理 3.6. 设 $\{A,B\}\in\mathbf{P}(m,p;n)$，则存在酉阵 $U\in\mathbf{C}^{m\times m}$ 和 $V\in\mathbf{C}^{p\times p}$，以及非奇异阵 $Q\in\mathbf{C}^{n\times n}$，使得

$$U^HAQ=\Sigma_A,\quad V^HBQ=\Sigma_B, \tag{3.29}$$

$$\Sigma_A=\begin{pmatrix}\Lambda & 0\\ 0 & 0_A\end{pmatrix}\begin{matrix}r+s\\ m-r-s,\end{matrix}$$
$$\qquad\quad\begin{matrix}r+s & n-r-s\end{matrix} \tag{3.30}$$

$$\Sigma_B=\begin{pmatrix}0_B & 0\\ 0 & \Omega\end{pmatrix}\begin{matrix}p+r-n\\ n-r,\end{matrix}$$
$$\qquad\quad\begin{matrix}r & n-r\end{matrix}$$

其中 0_A 与 0_B 为零矩阵，并且

$$\Lambda=\mathrm{diag}(\alpha_1,\cdots,\alpha_{r+s}),\Omega=\mathrm{diag}(\beta_{r+1},\cdots,\beta_n), \tag{3.31}$$

它们满足

$$\left.\begin{array}{l}1=\alpha_1=\cdots=\alpha_r>\alpha_{r+1}\geqslant\cdots\geqslant\alpha_{r+s}>\alpha_{r+s+1}\\ \quad=\cdots=\alpha_n=0\\ 0=\beta_1=\cdots=\beta_r<\beta_{r+1}\leqslant\cdots\leqslant\beta_{r+s}<\beta_{r+s+1}\\ \quad=\cdots=\beta_n=1\end{array}\right\} \tag{3.32}$$

和

$$\alpha_i^2+\beta_i^2=1,\quad i=1,\cdots,n. \tag{3.33}$$

作者证明了关于广义奇异值扰动的下列定理 3.7 和定理 3.8，它们分别是第三章定理 3.11 和定理 3.12 的推广. 定理 3.7 的证明，与定理 3.5 类似，见作者的论文[10]. 定理 3.8 的证明，见作者的论文[15].

定理 3.7[10]. 设 $\{A, B\} \in \mathbb{P}(m, p; n)$, $\{\widetilde{A}, \widetilde{B}\} = \{A + E, B + F\}$. 如果

$$\max_{\|x\|_2=1} \sqrt{\frac{\|Ex\|_2^2 + \|Fx\|_2^2}{\|Ax\|_2^2 + \|Bx\|_2^2}} < 1, \tag{3.34}$$

则 $\{\widetilde{A}, \widetilde{B}\} \in \mathbb{P}(m, p; n)$; 并且如果 $\{A, B\}$ 与 $\{\widetilde{A}, \widetilde{B}\}$ 的广义奇异值 $\{(\alpha_i, \beta_i)\}$ 与 $\{(\tilde{\alpha}_i, \tilde{\beta}_i)\}$ 都按 (3.32) 所示的顺序排列，则有

$$\rho((\alpha_i, \beta_i), (\tilde{\alpha}_i, \tilde{\beta}_i)) \leqslant t(\{A, B\}, \{\widetilde{A}, \widetilde{B}\}), i = 1, \cdots, n. \tag{3.35}$$

其中

$$\begin{aligned} & t(\{A, B\}, \{\widetilde{A}, \widetilde{B}\}) \\ & \quad \equiv \max_{\|x\|_2=1} \rho((\|Ax\|_2, \|Bx\|_2), (\|\widetilde{A}x\|_2, \|\widetilde{B}x\|_2)), \\ & \quad \{A, B\}, \{\widetilde{A}, \widetilde{B}\} \in \mathbb{P}(m, p; n). \end{aligned} \tag{3.36}$$

注 3.3 在 $\mathbb{P}(m, p; n)$ 中可以通过一种等价关系定义等价类，从而得到一个相应的投影空间，而 $t(\{A, B\}, \{\widetilde{A}, \widetilde{B}\})$ 恰是该投影空间上的一种投影度量.

注意到

$$\begin{aligned} t(\{A, B\}, \{\widetilde{A}, \widetilde{B}\}) & \leqslant \max_{\|x\|_2=1} \sqrt{\frac{\|Ex\|_2^2 + \|Fx\|_2^2}{\|Ax\|_2^2 + \|Bx\|_2^2}} \\ & \leqslant \left\{\sigma_{\max}\begin{pmatrix} E \\ F \end{pmatrix}\right\} \Big/ \left\{\sigma_{\min}\begin{pmatrix} A \\ B \end{pmatrix}\right\}, \end{aligned}$$

其中 $\sigma_{\max}(X)$ 与 $\sigma_{\min}(X)$ 分别表示 X 的最大与最小奇异值. 所以从定理 3.7 可得

推论 3.2. 设 $\{A, B\} \in \mathbb{P}(m, p; n)$, $\{\widetilde{A}, \widetilde{B}\} = \{A + E, B + F\}$. 如果

$$\sigma_{\max}\begin{pmatrix} E \\ F \end{pmatrix} < \sigma_{\min}\begin{pmatrix} A \\ B \end{pmatrix}, \tag{3.37}$$

则 $\{\widetilde{A}, \widetilde{B}\} \in \mathbb{P}(m, p; n)$; 并且对于 $\{A, B\}$ 与 $\{\widetilde{A}, \widetilde{B}\}$ 的广义奇异

值(均按(3.32)的顺序排列)，有

$$\rho((\alpha_i,\beta_i),(\tilde{\alpha}_i,\tilde{\beta}_i)) \leqslant \left\{\sigma_{\max}\begin{pmatrix} E \\ F \end{pmatrix}\right\} \Big/ \left\{\sigma_{\min}\begin{pmatrix} A \\ B \end{pmatrix}\right\},$$

$$i = 1, \cdots, n. \tag{3.38}$$

注 3.4 无妨假定推论 3,2 中 $\{A,B\}$ 与 $\{\tilde{A},\tilde{B}\}$ 的广义奇异值 (α_i,β_i) 与 $(\tilde{\alpha}_i,\tilde{\beta}_i)$ 满足

$$\alpha_i^2 + \beta_i^2 = \tilde{\alpha}_i^2 + \tilde{\beta}_i^2 = 1,\ i = 1, \cdots, n. \tag{3.39}$$

利用第二章定理 4.7，并注意到，对任一矩阵 X，有

$$\sigma_{\max}(X) = \|X\|_2,\quad \sigma_{\min}(X) = \|X^{\dagger}\|_2^{-1},$$

所以可把推论 3.2 的结论(3.38)改写成较弱、但较为直观的形式:

$$\sqrt{(\alpha_i - \tilde{\alpha}_i)^2 + (\beta_i - \tilde{\beta}_i)^2} \leqslant \sqrt{2}\left\|\begin{pmatrix} A \\ B \end{pmatrix}^{\dagger}\right\|_2 \left\|\begin{pmatrix} E \\ F \end{pmatrix}\right\|_2,$$

$$i = 1, \cdots, n. \tag{3.40}$$

定理 3.8[15]. 设 $\{A,B\}$ 与 $\{\tilde{A},\tilde{B}\} \in \mathbb{P}(m,p;n)$. $\sigma(A,B) = \{(\alpha_i,\beta_i)\}$ 与 $\sigma(\tilde{A},\tilde{B}) = \{(\tilde{\alpha}_i,\tilde{\beta}_i)\}$ 均按(3.32)排序. 如果令 $Z = \begin{pmatrix} A \\ B \end{pmatrix}$, $\tilde{Z} = \begin{pmatrix} \tilde{A} \\ \tilde{B} \end{pmatrix}$ 和

$$\rho_{i,i} = \rho((\alpha_i,\beta_i),(\tilde{\alpha}_i,\tilde{\beta}_i)),\quad i = 1, \cdots, n,$$

则有估计式

$$\prod_{i=1}^{n}(1 - \rho_{i,i}^2) \geqslant 1 - d_L^2(Z,\tilde{Z}). \tag{3.41}$$

上式右端的 $d_L(Z,\tilde{Z})$ (见第二章(4.34)和(4.26))为

$$d_L(Z,\tilde{Z}) \equiv \left\{1 - \frac{|\det Z^H\tilde{Z}|^2}{\det Z^H Z \det \tilde{Z}^H \tilde{Z}}\right\}^{\frac{1}{2}}.$$

注 3.5. 用类似于注 3.2 中的办法，可以分别从定理 3.7 和定理 3.8 推导出第三章的定理 3.11 和定理 3.12.

最近，Paige[134] 沿着作者 [15] 讨论广义奇异值扰动的途径，证明了下述定理.

定理 3.9[134]. 设 $\{A,B\}$ 与 $\{\tilde{A},\tilde{B}\} \in \mathbb{P}(m,p;n)$. $\sigma(A,B) = \{(\alpha_i,\beta_i)\}$ 与 $\sigma(\tilde{A},\tilde{B}) = \{(\tilde{\alpha}_i,\tilde{\beta}_i)\}$ 均按(3.32)排序，并假定

(α_i,β_i) 与 $(\tilde{\alpha}_i,\tilde{\beta}_i)$ 满足 (3.39). 如果令 $Z=\begin{pmatrix}A\\B\end{pmatrix}$ 和 $\tilde{Z}=\begin{pmatrix}\tilde{A}\\\tilde{B}\end{pmatrix}$，则有估计式

$$\sqrt{\sum_{i=1}^{n}[(\alpha_i-\tilde{\alpha}_i)^2+(\beta_i-\tilde{\beta}_i)^2]}\leqslant d_P(Z,\tilde{Z}), \quad (3.42)$$

其中 $d_P(Z,\tilde{Z})$ 由第二章 (4.31) 式定义.

证明:

令
$$Z_1= Z(Z^HZ)^{-\frac{1}{2}}=\begin{pmatrix}A_1\\B_1\end{pmatrix}\begin{matrix}m\\p\end{matrix},$$

$$\tilde{Z}_1= \tilde{Z}(\tilde{Z}^H\tilde{Z})^{-\frac{1}{2}}=\begin{pmatrix}\tilde{A}_1\\\tilde{B}_1\end{pmatrix}\begin{matrix}m\\p\end{matrix}. \quad (3.43)$$

根据定理 3.6，存在 $U,\tilde{U}\in\mathscr{U}_m$，$V$，$\tilde{V}\in\mathscr{U}_p$ 和 W，$\tilde{W}\in\mathscr{U}_n$，使得

$$\begin{pmatrix}U^H & 0\\0 & V^H\end{pmatrix}\begin{pmatrix}A_1\\B_1\end{pmatrix}W=\begin{pmatrix}\Sigma_A\\\Sigma_B\end{pmatrix},$$

$$\begin{pmatrix}\tilde{U}^H & 0\\0 & \tilde{V}^H\end{pmatrix}\begin{pmatrix}\tilde{A}_1\\\tilde{B}_1\end{pmatrix}\tilde{W}=\begin{pmatrix}\Sigma_{\tilde{A}}\\\Sigma_{\tilde{B}}\end{pmatrix}, \quad (3.44)$$

其中 Σ_A，Σ_B，$\Sigma_{\tilde{A}}$ 与 $\Sigma_{\tilde{B}}$ 如 (3.30)—(3.32) 所示. 由第三章定理 3.12 可知

$$\|\Sigma_A-\Sigma_{\tilde{A}}\|_F\leqslant\|A_1-\tilde{A}_1\|_F,\|\Sigma_B-\Sigma_{\tilde{B}}\|_F\leqslant\|B_1-\tilde{B}_1\|_F, \quad (3.45)$$

于是，有

$$\begin{aligned}&\sum_{i=1}^{n}[(\alpha_i-\tilde{\alpha}_i)^2+(\beta_i-\tilde{\beta}_i)^2]\\&\quad=\|\Sigma_A-\Sigma_{\tilde{A}}\|_F^2+\|\Sigma_B-\Sigma_{\tilde{B}}\|_F^2\\&\quad\leqslant\|A_1-\tilde{A}_1\|_F^2+\|B_1-\tilde{B}_1\|_F^2\\&\quad=\|Z_1-\tilde{Z}_1\|_F^2.\end{aligned} \quad (3.46)$$

注意到，用任一 $\tilde{Q}\in\mathscr{U}_n$ 从右端乘 $\tilde{Z}_1$ 并不改变不等式 (3.46)，所以从 (3.46) 立即得出

$$\sqrt{\sum_{i=1}^{n}[(\alpha_i-\tilde{\alpha}_i)^2+(\beta_i-\tilde{\beta}_i)^2]}\leqslant\min_{\tilde{Q}\in\mathscr{U}_n}\|Z_1-\tilde{Z}_1\tilde{Q}\|_F$$

$$= d_p(Z, \tilde{\tilde{Z}}).$$ □

注 3.6 根据第二章不等式(4.37)，有

$$d_p(Z, \tilde{Z}) \leqslant \sqrt{2}\, d_F(Z, \tilde{Z}),$$

其中 $d_F(Z, \tilde{Z})$ 由第二章(4.35)和(4.24)式定义，即

$$d_F(Z, \tilde{Z}) = \frac{1}{\sqrt{2}} \|P_Z - P_{\tilde{Z}}\|_F.$$

再利用本章§4的定理 4.6，可把定理 3.9 的结论 (3.42) 改写成较弱、但较为直观的形式:

$$\sqrt{\sum_{i=1}^{n} [(\alpha_i - \tilde{\alpha}_i)^2 + (\beta_i - \tilde{\beta}_i)^2]}$$

$$\leqslant \sqrt{2} \min\{\|Z^\dagger\|_2, \|\tilde{Z}^\dagger\|_2\} \left\| \begin{pmatrix} E \\ F \end{pmatrix} \right\|_F \tag{3.47}$$

$$\leqslant \sqrt{2} \left\| \begin{pmatrix} A \\ B \end{pmatrix}^\dagger \right\|_2 \left\| \begin{pmatrix} E \\ F \end{pmatrix} \right\|_F. \tag{3.48}$$

按照数值分析的观点，不等式(3.40)与(3.48)比起定理 3.7 —定理 3.9 的结论(见(3.35)、(3.41)和(3.42))更便于应用.

习题

1. 设 $A, B \in \mathbb{C}^{n\times n}$ 均为 Hermite 阵，并且 $B > 0$. 如果 $\lambda_1 \leqslant \cdots \leqslant \lambda_n$ 是 $\{A, B\}$ 的特征值，即存在非零向量 $\boldsymbol{x}_k \in \mathbb{C}^n$，使得 $A\boldsymbol{x}_k = \lambda_k B\boldsymbol{x}_k$, $k = 1, \cdots, n$，则有

$$\lambda_k = \min_{\substack{\mathscr{X} \subset \mathbb{C}^n \\ \dim(\mathscr{X}) = k}} \max_{\substack{\boldsymbol{x} \in \mathscr{X} \\ \boldsymbol{x} \neq 0}} \frac{\boldsymbol{x}^H A \boldsymbol{x}}{\boldsymbol{x}^H B \boldsymbol{x}}, \quad k = 1, \cdots, n$$

和

$$\lambda_k = \max_{\substack{\mathscr{X} \subset \mathbb{C}^n \\ \dim(\mathscr{X}) = n-k+1}} \min_{\substack{\boldsymbol{x} \in \mathscr{X} \\ \boldsymbol{x} \neq 0}} \frac{\boldsymbol{x}^H A \boldsymbol{x}}{\boldsymbol{x}^H B \boldsymbol{x}}, \quad k = 1, \cdots, n.$$

2. 设 $A, B \in \mathbb{C}^{n\times n}$ 均为 Hermite 阵，并且 $B > 0$. 又设 $\lambda(A) = \{\alpha_i\}$ 与 $\lambda(B) = \{\beta_i\}$ 满足 $\alpha_1 \geqslant \cdots \geqslant \alpha_n$ 与 $\beta_1 \geqslant \cdots \geqslant \beta_n$. 试证

$$\boldsymbol{x}^H B \boldsymbol{x} > |\boldsymbol{x}^H A \boldsymbol{x}| \quad \forall \boldsymbol{x} \in \mathbb{C}^n, \ \boldsymbol{x} \neq 0$$

的必要与充分条件是谱半径 $\rho(B^{-1}A) < 1$. 并且当 $\rho(B^{-1}A) < 1$

时，有

$$\beta_i > \alpha_i,\ \beta_i > -\alpha_{n-i+1},\ i = 1,\cdots,n.$$

3. 设 $\{A, B\} \in \mathrm{D}(n)$，$\hat{A}$ 与 $\hat{B}$ 分别是 A 与 B 相同位置上的 $n-1$ 阶主子阵. 试证：① $\{\hat{A},\hat{B}\} \in \mathrm{D}(n-1)$，$F(\hat{A}+i\hat{B}) \subseteq F(A+iB)$；②取一条从原点发出、不与 $A+iB$ 的值域 $F(A+iB)$ 相交的射线 R，利用 R 定义 $\{A, B\}$ 的特征角 $\theta_1 \leqslant \cdots \leqslant \theta_n$ 和 $\{\hat{A},\hat{B}\}$ 的特征角 $\hat{\theta}_1 \leqslant \cdots \leqslant \hat{\theta}_{n-1}$. 则有

$$\theta_1 \leqslant \hat{\theta}_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_{n-1} \leqslant \hat{\theta}_{n-1} \leqslant \theta_n.$$

4. 设 $\{A, B\} \in \mathrm{D}(n)$，$\lambda(A, B) = \{(\alpha_i, \beta_i)\}$. 任取 $\boldsymbol{x} \in \mathbf{C}^n$，$\|\boldsymbol{x}\|_2 = 1$. 令

$$r_A(\boldsymbol{x}) = \frac{\boldsymbol{x}^H A \boldsymbol{x}}{\sqrt{(\boldsymbol{x}^H A \boldsymbol{x})^2 + (\boldsymbol{x}^H B \boldsymbol{x})^2}},$$

$$r_B(\boldsymbol{x}) = \frac{\boldsymbol{x}^H B \boldsymbol{x}}{\sqrt{(\boldsymbol{x}^H A \boldsymbol{x})^2 + (\boldsymbol{x}^H B \boldsymbol{x})^2}}.$$

试证

$$\min_{1\leqslant i\leqslant n} \rho((r_A(\boldsymbol{x}), r_B(\boldsymbol{x})),\ (\alpha_i,\beta_i)) \leqslant \frac{\|(r_B(\boldsymbol{x})A - r_A(\boldsymbol{x})B)\boldsymbol{x}\|_2}{c(A,B)}.$$

并由此导出第三章定理 4.1.

5. 证明关于广义奇异值分解的定理 3.6.

§4 正规对、可对角化对与一般正则对的特征值

设 $\{A, B\}$ 是任一 n 阶正则对. 令 $Z = (A, B) \in \mathbf{C}^{n\times 2n}$. 由 $\det(A + \lambda B) \neq 0$ 容易推知 $\mathrm{rank}(Z) = n$. 此外，如果 $S \in \mathbf{C}_n^{n\times n}$，则 $\{SA, SB\}$ 亦为正则对，并且它与 $\{A, B\}$ 有相同的特征值与特征向量. 所以，就广义特征值问题来说，$\{A, B\}$ 与 $\{SA, SB\}$ 是等价的；换句话说，Z 与 SZ 是等价的，这里 $Z \in \mathbf{C}_n^{n\times 2n}$，$S \in \mathbf{C}_n^{n\times n}$. 因此，每一个正则对 $\{A, B\}$ 与 $\mathbf{G}_{n,2n}$ 上的一个点 $Z = (A, B)$ 一一对应. 本节就是从这个观点出发，把 Hoffman-Wielandt 定理、

Bauer-Fike 定理和 Henrici 定理推广到正则对的情形。

根据第二章 § 4 中的讨论，对于 $\mathbf{G}_{n,2n}$ 内的点 Z 与 W（这里的 Z 与 W 实际上是等价类中的代表元素），可以如下定义它们之间的距离，令

$$Z_1 = (ZZ^H)^{-\frac{1}{2}}Z, \quad W_1 = (WW^H)^{-\frac{1}{2}}W \tag{4.1}$$

和

$$\Theta(Z,W) \equiv \arccos(Z_1W_1^HW_1Z_1^H)^{\frac{1}{2}} \geqslant 0. \tag{4.2}$$

则

$$d_2(Z,W) = \|\sin\Theta(Z,W)\|_2 \tag{4.3}$$

与

$$d_F(Z,W) = \|\sin\Theta(Z,W)\|_F \tag{4.4}$$

是 $\mathbf{G}_{n,2n}$ 上的酉不变度量；并且有关系式

$$d_2(Z,W) = \|P_{Z^H} - P_{W^H}\|_2 \tag{4.5}$$

和

$$d_F(Z,W) = \frac{1}{\sqrt{2}}\|P_{Z^H} - P_{W^H}\|_F, \tag{4.6}$$

其中

$$P_{Z^H} = Z^\dagger Z = Z_1^HZ_1, \quad P_{W^H} = W^\dagger W = W_1^HW_1. \tag{4.7}$$

4.1 Hoffman-Wielandt 型定理

定理 4.1[9]. 设 $\{A, B\}$，$\{C, D\} \in \mathrm{N}(n)$，$\lambda(A,B) = \{(\alpha_i, \beta_i)\}$，$\lambda(C,D) = \{(\gamma_i,\delta_i)\}$. 令 $Z = (A, B)$，$W = (C, D)$ 以及

$$\rho_{i,j} = \rho((\alpha_i,\beta_i),(\gamma_j,\delta_j)), \ 1 \leqslant i,j \leqslant n. \tag{4.8}$$

则存在 $1,\cdots,n$ 的一个适当的排列 $\pi(1),\cdots,\pi(n)$，使得

$$\sqrt{\sum_{i=1}^{n}\rho_{i,\pi(i)}^2} \leqslant d_F(Z,W). \tag{4.9}$$

因而，有

$$\epsilon(\{A,B\},\{C,D\}) \leqslant d_F(Z,W). \tag{4.10}$$

证明:

由 (4.1)、(4.2) 和 (4.4) 可知

$$d_F^2(Z,W) = t_r[I-(AA^H+BB^H)^{-1}(AC^H+BD^H) \times (CC^H+DD^H)^{-1}(CA^H+DB^H)]. \tag{4.11}$$

根据定理 1.6，$\{A,B\}$ 与 $\{C,D\}$ 有分解式

$$A = R\Lambda U^H,\ B = R\Omega U^H,\ C = T\Gamma V^H,\ D = T\Delta V^H, \tag{4.12}$$

其中 R 与 $T\in\mathbf{C}^{n\times n}$ 为非奇异阵，U 与 $V\in\mathbf{C}^{n\times n}$ 为酉阵，

$$\left.\begin{aligned}\Lambda &= \mathrm{diag}(\alpha_1,\cdots,\alpha_n),\ \Omega = \mathrm{diag}(\beta_1,\cdots,\beta_n),\\ \Gamma &= \mathrm{diag}(\gamma_1,\cdots,\gamma_n),\ \Delta = \mathrm{diag}(\delta_1,\cdots,\delta_n).\end{aligned}\right\} \tag{4.13}$$

将 (4.12) 代入 (4.11) 式右端，并令 $W = U^H V$，可得

$$\begin{aligned} d_F^2(Z,W) &= \mathrm{tr}[I-(\Lambda\bar{\Lambda}+\Omega\bar{\Omega})^{-1}(\Lambda W\bar{\Gamma}+\Omega W\bar{\Delta}) \\ &\quad \times (\Gamma\bar{\Gamma}+\Delta\bar{\Delta})^{-1}(\Gamma W^H\bar{\Lambda}+\Delta W^H\bar{\Omega})] \\ &= n-g(W). \end{aligned} \tag{4.14}$$

其中 $W=(\omega_{ij})$ 为酉阵，并且

$$g(W) = \mathrm{tr}[(\Lambda\bar{\Lambda}+\Omega\bar{\Omega})^{-1}(\Lambda W\bar{\Gamma}+\Omega W\bar{\Delta}) \times (\Gamma\bar{\Gamma}+\Delta\bar{\Delta})^{-1}(\Gamma W^H\bar{\Lambda}+\Delta W^H\bar{\Omega})] \tag{4.15}$$

令

$$\theta_{ij} = \frac{|\alpha_i\bar{\gamma}_j+\beta_i\bar{\delta}_j|^2}{(|\alpha_i|^2+|\beta_i|^2)(|\gamma_j|^2+|\delta_j|^2)} \tag{4.16}$$

和

$$|\omega_{ij}|^2 = \sigma_{ij},\ 1\leqslant i,j\leqslant n, \tag{4.17}$$

则 $S=(\sigma_{ij})$ 显然是双随机阵，并且 (4.15) 可以写成

$$g(W) \equiv f(S) = \sum_{i,j=1}^{n}\theta_{ij}\sigma_{ij}. \tag{4.18}$$

用 $\mathscr{S}_n$ 表示 n 阶双随机阵的全体，用 $\mathscr{U}_n$ 表示 n 阶酉阵的全体。于是由 (4.15)—(4.18) 可知

$$\max_{W\in\mathscr{U}_n} g(W) \leqslant \max_{S\in\mathscr{S}_n} f(s). \tag{4.19}$$

利用 $f(S)$ 是 $\mathscr{S}_n$ 上的线性函数，并根据 Birkhoff 定理，可证（参看第三章 § 4 定理 4.2 的证明）：存在排列方阵

$$P=(\delta_{i,\pi(i)}),\quad \delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases},$$

使得

$$f(S)\leqslant f(P). \tag{4.20}$$

代入 (4.19) 和 (4.18)，得到

$$g(W)\leqslant\sum_{i=1}^{n}\theta_{i\pi(i)},$$

其中 $\pi(1),\cdots,\pi(n)$ 是 $1,\cdots,n$ 的某一个排列.

另一方面，若取 $W=P$，则有

$$g(P)=\sum_{i=1}^{n}\theta_{i\pi(i)};$$

所以

$$\max_{W\in\mathscr{U}_n}g(W)=\sum_{i=1}^{n}\theta_{i\pi(i)}$$

$$=\sum_{i=1}^{n}\frac{|\alpha_i\bar{\gamma}_{\pi(i)}+\beta_i\bar{\delta}_{\pi(i)}|^2}{(|\alpha_i|^2+|\beta_i|^2)(|\gamma_{\pi(i)}|^2+|\delta_{\pi(i)}|^2)}.$$

代入 (4.14)，便得出

$$d_F^2(Z,W)\geqslant\sum_{i=1}^{n}\rho_{i,\pi(i)}^2. \qquad \square$$

注 4.1. 同理可证，存在 $1,\cdots,n$ 的一个排列 $\sigma(1),\cdots,\sigma(n)$，使得

$$d_F(Z,W)\leqslant\sqrt{\sum_{i=1}^{n}\rho_{i,\sigma(i)}^2}.$$

注 4.2. 如果 $\{A,B\}$ 与 $\{C,D\}$ 不同时是正规对，则定理 4.1 的结论不一定成立. 例如

$$\{A,B\}=\left\{\begin{pmatrix}0&0\\0&-1\end{pmatrix},\begin{pmatrix}1&0\\0&1\end{pmatrix}\right\},$$

$$\{C,D\}=\left\{\begin{pmatrix}1&1\\0&1\end{pmatrix},\begin{pmatrix}0&1\\0&0\end{pmatrix}\right\}.$$

其中 $\{A,B\}$ 是正规对，而 $\{C,D\}$ 不是正规对. 容易看出

$$\lambda(A,B)=\{(0,1),(-1,1)\},$$
$$\lambda(C,D)=\{(1,0),(1,0)\}.$$

有

$$\rho_{1,j}^2=1,\quad \rho_{2,j}^2=\frac{1}{2},\quad 1\leqslant j\leqslant 2.$$

不管如何选取 1，2 的排列 $\pi(1)$，$\pi(2)$，总有

$$\sum_{i=1}^{2}\rho_{i,\pi(i)}^2=\frac{3}{2}>\frac{5}{4}=d_F^2(Z,W),$$

其中 $Z=(A,B)$，$W=(C,D)$.

注 4.3. 由定理 4.1 可以导出 Hoffman-Wielandt 定理(第三章定理 4.2). 事实上,如果 $A,C\in\mathbf{C}^{n\times n}$ 均为正规阵,并且 $\lambda(A)=\{\alpha_i\}$ 和 $\lambda(C)=\{\gamma_i\}$，则对于 $t>0$，$\{A,\ tI\}$ 与 $\{C,\ tI\}\in\mathbf{D}(n)$，并且有 $\lambda(A,tI)=\{(\alpha_i,t)\}$ 和 $\lambda(C,tI)=\{(\gamma_i,t)\}$. 令 $Z_t=(A,tI)$ 和 $W_t=(C,tI)$. 于是根据定理 4.1，存在 $1,\cdots,n$ 的某一个固定的排列 $\pi(1),\cdots,\pi(n)$ 和一个趋于 $+\infty$ 的正数序列 $t_1,t_2,\cdots$，使得

$$\sum_{i=1}^{n}\rho^2((\alpha_i,t_j),(\gamma_{\pi(i)},t_j))\leqslant d_F^2(Z_{t_j},W_{t_j}),$$
$$j=1,2,\cdots,$$

即

$$\sum_{i=1}^{n}\frac{|\alpha_i-\gamma_{\pi(i)}|^2t_j^2}{(|\alpha_i|^2+t_j^2)(|\gamma_{\pi(i)}|^2+t_j^2)}$$
$$\leqslant \mathrm{tr}[I-(I+A_{t_j}A_{t_j}^H)^{-1}(I+A_{t_j}C_{t_j}^H)(I+C_{t_j}C_{t_j}^H)^{-1}$$
$$(I+C_{t_j}A_{t_j}^H)],\tag{4.21}$$

其中 $A_{t_j}=\frac{1}{t_j}A$, $C_{t_j}=\frac{1}{t_j}C$. 当 t_j 足够大时,将(4.21)式右端的 $(I+A_{t_j}A_{t_j}^H)^{-1}$ 和 $(I+C_{t_j}C_{t_j}^H)^{-1}$ 展开,从而可得

$$\sum_{i=1}^{n}|\alpha_i-\gamma_{\pi(i)}|^2\leqslant\|A-C\|_F^2+O\left(\frac{1}{t_j^2}\right)\ (t_j\to+\infty).$$

令 $t_j\to+\infty$，便给出第三章 § 4 的不等式 (4.2).

下面给出正规对的一个判别定理.

定理 4.2. 设 $\{A,B\}$ 为 n 阶正则对. 令

$$(A_1,B_1)=(AA^H+BB^H)^{-\frac{1}{2}}(A,B). \tag{4.22}$$

则 $\{A, B\}\in \mathrm{N}(n)$ 的必要与充分条件是 $A_1B_1^H$ 和 $B_1{}^HA_1$ 均为正规阵.

证明:

充分性. 设 $A_1B_1^H$ 和 $B_1^HA_1$ 均为正规阵,则根据 Williamson [194](见 § 1 习题 4),存在酉阵 U 与 V,使得

$$A_1=V\mathrm{diag}(\alpha_1,\cdots,\alpha_n)U^H,$$
$$B_1=V\mathrm{diag}(\beta_1,\cdots,\beta_n)U^H.$$

代入 (4.22) 式,令 $(AA^H+BB^H)^{-\frac{1}{2}}V=S$, 则 $\{A,B\}$有分解式

$$S^HAU=\mathrm{diag}(\alpha_1,\cdots,\alpha_n),$$
$$S^HBU=\mathrm{diag}(\beta_1,\cdots,\beta_n). \tag{4.23}$$

根据定理 1.6, 立即得知 $\{A,B\}\in \mathrm{N}(n)$.

必要性. 设 $\{A,B\}\in \mathrm{N}(n)$, 则据定理 1.6, $\{A,B\}$ 有(4.23)式所示的分解,其中 S 为非奇异阵,U 为酉阵. 令

$$\Lambda=\mathrm{diag}(\alpha_1,\cdots,\alpha_n),\ \ \Omega=\mathrm{diag}(\beta_1,\cdots,\beta_n),$$

则有

$$A_1=(AA^H+BB^H)^{-\frac{1}{2}}A=[S(\Lambda\bar{\Lambda}+\Omega\bar{\Omega})^{-1}S^H]^{\frac{1}{2}}S^{-H}\Lambda U^H$$

和

$$B_1=(AA^H+BB^H)^{-\frac{1}{2}}B=[S(\Lambda\bar{\Lambda}+\Omega\bar{\Omega})^{-1}S^H]^{\frac{1}{2}}S^{-H}\Omega U^H.$$

令

$$W_1=[S(\Lambda\bar{\Lambda}+\Omega\bar{\Omega})^{-1}S^H]^{\frac{1}{2}}S^{-H}(\Lambda\bar{\Lambda}+\Omega\bar{\Omega})^{\frac{1}{2}}$$

和

$$\Lambda_1=(\Lambda\bar{\Lambda}+\Omega\bar{\Omega})^{-1}\Lambda\bar{\Omega},$$

计算可知: W_1 是酉阵,并且有

$$A_1B_1^H=W_1\Lambda_1W_1^H,\ \ B_1^HA_1=U\Lambda_1U^H.$$

即 $A_1B_1^H$ 与 $B_1^HA_1$ 均为正规阵. □

利用定理 4.2, 可以判别注 4.2 中的 $\{C,D\}\notin \mathrm{N}(2)$. 因为

$$D_1^HC_1=D^H(CC^H+DD^H)^{-1}C$$

$$= \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ \frac{1}{2} & 0 \end{pmatrix}$$

显然不是正规阵.

4.2 Bauer-Fike 型定理

定理 4.3[9][187]. 设 $\{A,B\} \in \mathbf{D}_g(n)$ 有分解式

$$A = P\mathrm{diag}(\alpha_1, \cdots, \alpha_n)Q, \quad B = P\mathrm{diag}(\beta_1, \cdots, \beta_n)Q, \tag{4.24}$$

其中 P 与 $Q \in \mathbb{C}^{n\times n}$ 为非奇异阵. 又设 $\{C, D\}$ 为 n 阶正则对. 令 $Z = (A,B)$ 和 $W = (C,D)$. 则

$$s_{\{A,B\}}\{C,D\} \leqslant \|Q\|_2 \|Q^{-1}\|_2 d_2(Z,W). \tag{4.25}$$

$s_{\{A,B\}}\{C,D\}$ 与 $d_2(Z,W)$ 分别如 (1.40) 和 (4.3) 式所示.

证明:

令

$$Z_1 = (ZZ^H)^{-\frac{1}{2}}Z \equiv (A_1, B_1),$$
$$W_1 = (WW^H)^{-\frac{1}{2}}W \equiv (C_1, D_1). \tag{4.26}$$

根据 (4.24), $\{A_1, B_1\}$ 有分解式

$$A_1 = P_1 \Lambda Q, \quad B_1 = P_1 \Omega Q,$$

其中 P_1 为非奇异阵, $\Lambda = \mathrm{diag}(\alpha_1, \cdots, \alpha_n), \Omega = \mathrm{diag}(\beta_1, \cdots, \beta_n)$.

任取 $(\gamma, \delta) \in \lambda(C,D) = \lambda(C_1, D_1)$, 无妨假设 $|\gamma|^2 + |\delta|^2 = 1$. 设 $\boldsymbol{x} \in \mathbb{C}^n$ 是 $\{C,D\}$ 属于 (γ,δ) 的单位特征向量,即

$$\delta C_1 \boldsymbol{x} = \gamma D_1 \boldsymbol{x}, \quad \|\boldsymbol{x}\|_2 = 1.$$

一方面, 有

$$\delta A_1 \boldsymbol{x} - \gamma B_1 \boldsymbol{x} = \delta(A_1 - Z_1 W_1^H C_1)\boldsymbol{x} - \gamma(B_1 - Z_1 W_1^H D_1)\boldsymbol{x}$$

$$= (A_1 - Z_1 W_1^H C_1, B_1 - Z_1 W_1^H D_1) \begin{pmatrix} \delta \boldsymbol{x} \\ -\gamma \boldsymbol{x} \end{pmatrix}$$

$$= (Z_1 - Z_1 W_1^H W_1)\begin{pmatrix}\delta \boldsymbol{x}\\ -\gamma \boldsymbol{x}\end{pmatrix}$$

$$= Z_1(Z_1^H Z_1 - W_1^H W_1)\begin{pmatrix}\delta \boldsymbol{x}\\ -\gamma \boldsymbol{x}\end{pmatrix}. \tag{4.27}$$

于是利用 (4.5), (4.7), $|\gamma|^2 + |\delta|^2 = 1$, $\|\boldsymbol{x}\|_2 = 1$ 以及 $\|Z_1\|_2 = 1$，得到

$$\|\delta A_1 \boldsymbol{x} - \gamma B_1 \boldsymbol{x}\|_2 \leqslant \|Z_1^H Z_1 - W_1^H W_1\|_2 = d_2(Z, W). \tag{4.28}$$

另一方面，利用

$$I = Z_1 Z_1^H = P_1(\Lambda Q Q^H \bar{\Lambda} + \Omega Q Q^H \bar{\Omega}) P_1^H$$

可得

$$P_1^H P_1 = (\Lambda Q Q^H \bar{\Lambda} + \Omega Q Q^H \bar{\Omega})^{-1}$$

$$\geqslant \frac{1}{\|Q\|_2^2}(\Lambda\bar{\Lambda} + \Omega\bar{\Omega})^{-1},$$

从而

$$\|\delta A_1 \boldsymbol{x} - \gamma B_1 \boldsymbol{x}\|_2 = \|P_1(\delta\Lambda - \gamma\Omega) Q \boldsymbol{x}\|_2$$

$$= [\boldsymbol{x}^H Q^H (\delta\Lambda - \gamma\Omega)^H P_1^H P_1 (\delta\Lambda - \gamma\Omega) Q \boldsymbol{x}]^{\frac{1}{2}}$$

$$\geqslant \frac{1}{\|Q\|_2}[\boldsymbol{x}^H Q^H (\delta\Lambda - \gamma\Omega)^H (\Lambda\bar{\Lambda} + \Omega\bar{\Omega})^{-1}$$

$$\times (\delta\Lambda - \gamma\Omega) Q \boldsymbol{x}]^{\frac{1}{2}}$$

$$\geqslant \frac{1}{\|Q\|_2}\min_{1 \leqslant i \leqslant n} \frac{|\delta\alpha_i - \gamma\beta_i|}{\sqrt{|\alpha_i|^2 + |\beta_i|^2}} \cdot (\boldsymbol{x}^H Q^H Q \boldsymbol{x})^{\frac{1}{2}}$$

$$\geqslant \frac{1}{\|Q\|_2 \|Q^{-1}\|_2}\min_{1 \leqslant i \leqslant n} \rho((\alpha_i, \beta_i), (\gamma, \delta)). \tag{4.29}$$

把不等式(4.28)和(4.29)联系起来，便得到估计式(4.25). □

注 4.4. 由定理 4.3 可以导出 Bauer-Fike 定理（第三章定理 4.5）. 事实上，如果 $A, C \in \mathbf{C}^{n \times n}$，其中 A 为可对角化阵，

$$A = Q^{-1} \Lambda Q, \quad \Lambda = \mathrm{diag}(\alpha_1, \cdots, \alpha_n).$$

则对于 $t > 0$，

$$\{A, tI\} \in \mathrm{D}_g(n), \quad \lambda(A, tI) = \{(\alpha_i, t)\}, \quad \{C, tI\}$$

为 n 阶正则对. 令 $Z_t = (A, tI)$ 和 $W_t = (C, tI)$. 于是根据定

理 4.3，对于任一 $\gamma\in\lambda(C)$，存在一个固定的指标 $i\,(1\leqslant i\leqslant n)$ 和一个趋于 $+\infty$ 的正数序列 $t_1,t_2,\cdots$，使得

$$\rho^2((\alpha_i,t_j),(\gamma,t_j))\leqslant d_2^2(Z_{t_j},W_{t_j}),$$

即

$$\begin{aligned}&\frac{|\alpha_i-\gamma|^2t_j^2}{(|\alpha_i|^2+t_j^2)(|\gamma|^2+t_j^2)}\\&\quad\leqslant\|I-(Z_{t_j}Z_{t_j}^H)^{-\frac12}Z_{t_j}W_{t_j}^H(W_{t_j}W_{t_j}^H)^{-1}W_{t_j}Z_{t_j}^H(Z_{t_j}Z_{t_j}^H)^{-\frac12}\|_2\\&\quad=\lambda_{\max}\left[I-\left(I+\frac{1}{t_j^2}AA^H\right)^{-1}\left(I+\frac{1}{t_j^2}AC^H\right)\right.\\&\qquad\left.\times\left(I+\frac{1}{t_j^2}CC^H\right)^{-1}\left(I+\frac{1}{t_j^2}CA^H\right)\right].\end{aligned}$$

当 t_j 足够大时，可将

$$\left(I+\frac{1}{t_j^2}AA^H\right)^{-1}\text{ 和 }\left(I+\frac{1}{t_j^2}CC^H\right)^{-1}$$

展开，于是得到

$$|\alpha_i-\gamma|^2\leqslant\|A-C\|_2^2+O\left(\frac{1}{t_j^2}\right)(t_j\to+\infty).$$

令 $t_j\to+\infty$，便得出第三章定理 4.5 的结论。

从定理 4.3 可以立即得到

推论4.1. 设 $\{A,B\}\in\mathbb{N}(n)$，$\{C,D\}$ 为任一 n 阶正则对。令 $Z=(A,B)$ 和 $W=(C,D)$。则

$$s_{\{A,B\}}\{C,D\}\leqslant d_2(Z,W).\tag{4.30}$$

此外，还可以导出

推论 4.2. 设 $\{A,B\}\in\mathbb{D}(n)$，$\{C,D\}$ 为任一 n 阶正则对。令 $Z=(A,B)$ 和 $W=(C,D)$。则

$$s_{\{A,B\}}\{C,D\}\leqslant\frac{\|Z\|_2}{c(A,B)}d_2(Z,W).\tag{4.31}$$

证明:

根据定理 1.8，存在非奇异阵 $Q\in\mathbb{C}^{n\times n}$，使得

$$Q^HAQ=\Lambda,\ Q^HBQ=\Omega,\tag{4.32}$$

其中

$$\Lambda = \mathrm{diag}(\alpha_1, \cdots, \alpha_n), \quad \Omega = \mathrm{diag}(\beta_1, \cdots, \beta_n),$$

并且无妨假设 $\alpha_i^2 + \beta_i^2 = 1, i=1,\cdots,n$. 于是,对于任一 $\boldsymbol{x} \in \mathbb{C}^n$, $\|\boldsymbol{x}\|_2 = 1$, 有

$$\frac{|\boldsymbol{x}^H(\Lambda + i\Omega)\boldsymbol{x}|}{\boldsymbol{x}^H Q^H Q \boldsymbol{x}} = \frac{|\boldsymbol{x}^H Q^H (A + iB) Q \boldsymbol{x}|}{\boldsymbol{x}^H Q^H Q \boldsymbol{x}} \geqslant c(A, B),$$

其中 $c(A,B)$ 按 (1.47) 式定义. 因此,

$$\begin{aligned} \|Q\boldsymbol{x}\|_2^2 &\leqslant \frac{1}{c(A,B)} |\boldsymbol{x}^H(\Lambda + i\Omega)\boldsymbol{x}| \\ &\leqslant \frac{1}{c(A,B)}, \quad \forall \boldsymbol{x} \in \mathbb{C}^n, \ \|\boldsymbol{x}\|_2 = 1. \end{aligned}$$

由此可知

$$\|Q\|_2 \leqslant \frac{1}{\sqrt{c(A,B)}}. \tag{4.33}$$

再者,根据 (4.32) 和 $\Lambda^2 + \Omega^2 = I$, 有

$$I = Q^H(AQQ^HA + BQQ^HB)Q,$$

因而

$$\|Q^{-1}\|_2 \leqslant \|Q\|_2 \|A^2 + B^2\|_2^{1/2} \leqslant \frac{\|Z\|_2}{\sqrt{c(A,B)}}. \tag{4.34}$$

将(4.33)和(4.34)代入 (4.25) 式右端, 便导出不等式 (4.31). □

4.3 Henrici 型定理

设 $\{A,B\}$ 是 n 阶正则对. 令 $\mathscr{M}_{\{A,B\}}$ 表示所有满足下述三个条件的矩阵对 $\{X,U\}$ 的集合:

$$X, U \in \mathbb{C}^{n\times n}, \ X \text{ 是非奇异阵}, U \text{ 是酉阵}; \tag{4.35a}$$

$$XAU \text{ 与 } XBU \text{ 均为上三角阵}; \tag{4.35b}$$

$$|(XAU)_{ii}|^2 + |(XBU)_{ii}|^2 = 1, \ i = 1, \cdots, n. \tag{4.35c}$$

此处 $(A)_{ii}$ 表示矩阵 A 的第 (i,i) 位置的元素.

由定理 1.3 知, $\mathscr{M}_{\{A,B\}} \neq \varnothing$.

设 $\{X,U\} \in \mathscr{M}_{\{A,B\}}$. 用 $\mathrm{diag}(A)$ 表示由 A 的对角线元素组

成的对角阵．记

$$\mu(X,U)=\|(XAU-\mathrm{diag}(XAU),\ \ XBU-\mathrm{diag}(XBU))\|_2. \tag{4.36}$$

令

$$\Delta_2(A,B)=\inf_{\{X,U\}\in\mathscr{M}_{\{A,B\}}}\mu(X,U). \tag{4.37}$$

$\Delta_2(A,B)$ 叫做 $\{A,B\}$ 的正规性偏离度．

关于 $\Delta_2(A,B)$，需要指出两点：

1）(4.37) 式所示的下确界，可以在 $\mathscr{M}_{\{A,B\}}$ 上达到．

证明：

任取一固定的 $\{X^{(0)},U^{(0)}\}\in\mathscr{M}_{\{A,B\}}$．设

$$X^{(0)}AU^{(0)}=D_A^{(0)}+M_A^{(0)},\quad X^{(0)}BU^{(0)}=D_B^{(0)}+M_B^{(0)},$$

其中 $D_A^{(0)}$ 与 $D_B^{(0)}$ 为对角阵，满足条件 (4.35c)，$M_A^{(0)}$ 与 $M_B^{(0)}$ 为严格上三角阵．与此同时，取一固定的 $\lambda_0\in\mathbf{C}$，使

$$\det(A+\lambda_0 B)\neq 0,$$

于是有

$$X^{(0)}(A+\lambda_0 B)U^{(0)}=D_A^{(0)}+M_A^{(0)}+\lambda_0(D_B^{(0)}+M_B^{(0)}),$$

从而得出

$$X^{(0)}=[(D_A^{(0)}+\lambda_0 D_B^{(0)})+(M_A^{(0)}+\lambda_0 M_B^{(0)})]U^{(0)-1}(A+\lambda_0 B)^{-1}$$

和

$$\begin{aligned}\|X^{(0)}\|_2&\leqslant[\|D_A^{(0)}+\lambda_0 D_B^{(0)}\|_2+\|M_A^{(0)}+\lambda_0 M_B^{(0)}\|_2]\|(A+\lambda_0 B)^{-1}\|_2\\&\leqslant\sqrt{1+|\lambda_0|^2}(1+\|(M_A^{(0)},M_B^{(0)})\|_2)\|(A+\lambda_0 B)^{-1}\|_2\\&=\sqrt{1+|\lambda_0|^2}(1+\mu(X^{(0)},U^{(0)}))\|(A+\lambda_0 B)^{-1}\|_2\\&\equiv\gamma_0.\end{aligned} \tag{4.38}$$

由 (4.38) 可知，如果 $\{X,U\}\in\mathscr{M}_{\{A,B\}}$，使得

$$\mu(X,U)\leqslant\mu(X^{(0)},U^{(0)}),$$

则必有 $\|X\|_2\leqslant\gamma_0$．所以根据 (4.37) 式，若令

$$\mathscr{M}_{\{A,B\}}^{(0)}=\{\{X,U\}\in\mathscr{M}_{\{A,B\}}:\|X\|_2\leqslant\gamma_0\}, \tag{4.39}$$

则

$$\Delta_2(A,B)=\inf_{\{X,U\}\in\mathscr{M}_{\{A,B\}}^{(0)}}\mu(X,U). \tag{4.40}$$

再注意到 $\mathscr{M}^{(0)}_{\{A,B\}}$ 是 $\mathbf{C}^{2n^2}$ 内的有界闭集（(4.39) 式中的 $\|X\|_2 \leqslant \gamma_0$ 和 $\|U\|_2 \leqslant 1$ 已说明 $\mathscr{M}^{(0)}_{\{A,B\}}$ 是一有界集；此外，容易看出，如果 $\{X, U\}$ 是 $\mathscr{M}^{(0)}_{\{A,B\}}$ 内一列点集的极限点，则由 (4.28)—(4.30) 可知，U 必为酉阵，X 必为非奇异阵，并且 X 与 U 满足条件 (4.35a)—(4.35c) 和 (4.38)，即 $\{X,U\} \in \mathscr{M}^{(0)}_{\{A,B\}}$，因而 $\mathscr{M}^{(0)}_{\{A,B\}}$ 为闭集）. 因此，作为 $\mathscr{M}^{(0)}_{\{A,B\}}$ 上的连续函数的 $\mu(X, U)$，必在 $\mathscr{M}^{(0)}_{\{A,B\}}$ 上达到极小值. □

因此，定义式 (4.37) 可以改写为

$$\Delta_2(A,B) = \min_{\{X,U\} \in \mathscr{M}_{\{A,B\}}} \mu(X,U). \tag{4.41}$$

并且由此可以立即得到

2) 设 $\{A,B\}$ 为 n 阶正则对，则

$$\{A,B\} \in \mathrm{N}(n) \Longleftrightarrow \Delta_2(A,B) = 0. \tag{4.42}$$

定理 4.4[87]. 设 $\{A,B\}$ 与 $\{C,D\}$ 均为 n 阶正则对. $\Delta_2(A,B)$ 表示 $\{A, B\}$ 的正规性偏离度，并且假定 $\Delta_2(A, B) \neq 0$. 令 $Z = (A,B)$ 和 $W = (C,D)$. 则

$$s_{\{A,B\}}\{C,D\} \leqslant \frac{\eta}{g(\eta)}(1 + \Delta_2(A,B))d_2(Z,W). \tag{4.43}$$

其中

$$\eta = \frac{\Delta_2(A,B)}{[1 + \Delta_2(A,B)]d_2(Z,W)}, \tag{4.44}$$

$g(\eta)$ 是 $g + g^2 + \cdots + g^n = \eta(\eta > 0)$ 的唯一非负解.

证明:

首先注意到，利用一个 n 阶非奇异阵同时从左边乘以 A 和 B，并不改变 $s_{\{A,B\}}\{C,D\}$，$\Delta_2(A,B)$ 和 (4.44) 所示的 η 之值，所以无妨假定

$$A = (D_A + M_A)U^H, \quad B = (D_B + M_B)U^H. \tag{4.45}$$

其中 U 为酉阵； $D_A = \mathrm{diag}(\alpha_1, \cdots, \alpha_n)$，$D_B = \mathrm{diag}(\beta_1, \cdots, \beta_n)$，满足 $|\alpha_i|^2 + |\beta_i|^2 = 1$，$i = 1, \cdots, n$；$M_A$ 与 M_B 为严格上三角阵，并且 $\|(M_A, M_B)\|_2 = \Delta_2(A,B)$.

任取一 $(\gamma,\delta) \in \lambda(C, D)$. 显然只需考虑 $(\gamma, \delta) \notin \lambda(A, B)$

的情形．无妨假设 $|\gamma|^2+|\delta|^2=1$，并且 $\delta C\boldsymbol{x}=\gamma D\boldsymbol{x}$，$\boldsymbol{x}\in\mathbf{C}^n$，$\|\boldsymbol{x}\|_2=1$．令

$$
\begin{aligned}
R &= \delta(D_A+M_A)-\gamma(D_B+M_B)\\
&= (\delta D_A-\gamma D_B)+(\delta M_A-\gamma M_B).
\end{aligned}
$$

由 $RU=\delta A-\gamma B$ 可知 R 为非奇异阵，并且有估计式

$$
\begin{aligned}
\|R^{-1}\|_2^{-1} &= \|(RU)^{-1}\|_2^{-1}\\
&= \|(\delta A-\gamma B)^{-1}\|_2^{-1}\leqslant\|\delta A\boldsymbol{x}-\gamma B\boldsymbol{x}\|_2.
\end{aligned}
$$

利用关系式 (4.26) 和 (4.27)，有

$$
\delta A\boldsymbol{x}-\gamma B\boldsymbol{x}=Z(P_{Z^H}-P_{W^H})\begin{pmatrix}\delta\boldsymbol{x}\\-\gamma\boldsymbol{x}\end{pmatrix},
$$

因此，

$$
\begin{aligned}
\|R^{-1}\|_2^{-1} &\leqslant \|Z\|_2 d_2(Z,W)\\
&\leqslant [\|(D_A,D_B)\|_2+\|(M_A,M_B)\|_2]d_2(Z,W)\\
&= [1+\Delta_2(A,B)]d_2(Z,W)\equiv\varepsilon\ . \qquad (4.46)
\end{aligned}
$$

再注意到 R 的严格上三角部分 $\delta M_A-\gamma M_B$ 满足

$$
\|\delta M_A-\gamma M_B\|_2\leqslant\|(M_A,M_B)\|_2=\Delta_2(A,B)\neq 0.
$$

所以根据第三章引理 5.1 (只需将该引理中的条件改为 $\|M\|_2\leqslant m\neq 0$ 即可)，可得

$$
\min_{1\leqslant i\leqslant n}|\delta\alpha_i-\gamma\beta_i|\leqslant\frac{\Delta_2(A,B)}{g(\Delta_2(A,B)/\varepsilon)}.
$$

将 $|\delta\alpha_i-\gamma\beta_i|=\rho((\alpha_i,\beta_i),(\gamma,\delta))$ 代入上式左端，便导出了估计式 (4.43)．□

注 4.5． 当 $\{A,B\}\in\mathrm{N}(n)$ 时，有 $\Delta_2(A,B)=0$．因此，利用函数 $g(\eta)$ 的性质(见第三章 (5.15) 式)

$$
\lim_{\eta\to+0}\frac{\eta}{g(\eta)}=1,
$$

从估计式(4.43)可推导出估计式(4.30)．

注 4.6． 由定理 4.4 可以导出 Henrici 定理(第三章定理 5.2)．这可以仿照注 4.4 所述的办法进行推导，只是需要指出一点，即：如果 $A\in\mathbf{C}^{n\times n}$ 有分解式 $U^HAU=D+M$，其中 U 为酉阵，

$D = \mathrm{diag}(\alpha_1, \cdots, \alpha_n)$，$M$ 为严格上三角阵；则对于 $t > 0$，根据条件 (4.35a)—(4.35c)，应将正则对 $\{A, tI\}$ 分解为

$$XAU = D_A + M_A, \quad X(tI)U = D_B,$$

其中

$$X = \Lambda U^H, \ D_A = \Lambda D, \ M_A = \Lambda M, \ D_B = t\Lambda,$$

$$\Lambda = \mathrm{diag}\left(\frac{1}{\sqrt{|\alpha_1|^2 + t^2}}, \cdots, \frac{1}{\sqrt{|\alpha_n|^2 + t^2}}\right).$$

因此，

$$\begin{aligned}\Delta_2(A, tI) &\leqslant \|M_A\|_2 = \|\Lambda M\|_2 \\ &\leqslant \|\Lambda\|_2 \|M\|_2 \leqslant \frac{\|M\|_2}{t},\end{aligned}$$

从而有

$$\Delta_2(A, tI) \leqslant \frac{\Delta_2(A)}{t}.$$

$\Delta_2(A)$ 表示 A 对于谱范数 $\|\ \ \|_2$ 的正规性偏离度（见第三章定义 5.1）.

4.4 $d_2(Z,W)$ 与 $d_F(Z,W)$ 的估计

本节关于广义特征值所得到的扰动界限中，都使用了投影度量 $d_2(Z,W)$ 或 $d_F(Z,W)$. 我们在这段里，将研究这两个投影度量与 $W-Z$ 的关系，并给出 $d_2(Z,W)$ 与 $d_F(Z,W)$ 的上、下界.

定理 4.5. 设 $Z, W \in \mathbf{C}_n^{n\times 2n}$. 令 $P_{Z^H}^{\perp} = I - P_{Z^H}$，$P_{W^H}^{\perp} = I - P_{W^H}$. 则

$$d_p(Z, W) = \|(WW^H)^{-\frac{1}{2}}(W - Z)P_{Z^H}^{\perp}\|_p \tag{4.47}$$

$$= \|(ZZ^H)^{-\frac{1}{2}}(W - Z)P_{W^H}^{\perp}\|_p, \tag{4.48}$$

其中 $p = 2, F$.

证明:

分解

$$Z = (\underset{n}{Z_1}, \ \underset{n}{0})U, \quad W = (\underset{n}{W_1}, \ \underset{n}{W_2})U,$$

其中 $U\in\mathbf{C}^{2n\times 2n}$ 为酉阵. 可算出

$$P_{W^H}-P_{Z^H}=W^H(WW^H)^{-1}W-Z^H(ZZ^H)^{-1}Z=U^H$$

$$\times\begin{pmatrix} W_1^H(W_1W_1^H+W_2W_2^H)^{-1}W_1-I & W_1^H(W_1W_1^H+W_2W_2^H)^{-1}W_2 \\ W_2^H(W_1W_1^H+W_2W_2^H)^{-1}W_1 & W_2^H(W_1W_1^H+W_2W_2^H)^{-1}W_2 \end{pmatrix}U$$

和

$$(P_{W^H}-P_{Z^H})^2=U^H$$

$$\times\begin{pmatrix} I-W_1^H(W_1W_1^H+W_2W_2^H)^{-1}W_1 & 0 \\ 0 & W_2^H(W_1W_1^H+W_2W_2^H)^{-1}W_2 \end{pmatrix}U. \tag{4.49}$$

另一方面, 由

$$P_{Z^H}^{\perp}=U^H\begin{pmatrix} 0 & 0 \\ 0 & I^{(n)} \end{pmatrix}U$$

可知

$$[(W-Z)P_{Z^H}^{\perp}]^H(WW^H)^{-1}[(W-Z)P_{Z^H}^{\perp}]$$

$$=U^H\begin{pmatrix} 0 & 0 \\ 0 & W_2^H(W_1W_1^H+W_2W_2^H)^{-1}W_2 \end{pmatrix}U. \tag{4.50}$$

注意到

$$\begin{aligned}
&\lambda(W_2^H(W_1W_1^H+W_2W_2^H)^{-1}W_2) \\
&\quad=\lambda((W_1W_1^H+W_2W_2^H)^{-1}W_2W_2^H) \\
&\quad=\lambda(I-(W_1W_1^H+W_2W_2^H)^{-1}W_1W_1^H) \\
&\quad=\lambda(I-W_1^H(W_1W_1^H+W_2W_2^H)^{-1}W_1),
\end{aligned}$$

所以由 (4.49) 和 (4.50) 可知,如果 $(WW^H)^{-\frac{1}{2}}[(W-Z)P_{Z^H}^{\perp}]$ 的奇异值为 $\sigma_1,\cdots,\sigma_n$, 则 $P_{W^H}-P_{Z^H}$ 的奇异值为 $\sigma_1,\sigma_1,\cdots,\sigma_n,\sigma_n$. 因此,有

$$\|P_{W^H}-P_{Z^H}\|_2=\|(WW^H)^{-\frac{1}{2}}(W-Z)P_{Z^H}^{\perp}\|_2$$

和

$$\frac{1}{\sqrt{2}}\|P_{W^H}-P_{Z^H}\|_F=\|(WW^H)^{-\frac{1}{2}}(W-Z)P_{Z^H}^{\perp}\|_F.$$

即等式 (4.47) 成立. 同理可证等式 (4.48). □

从 (4.47) 和 (4.48) 立即得到

定理 4.6. 设 $Z, W \in \mathbf{C}_n^{n\times 2n}$. 则有估计式

$$\max\left\{\frac{\|(W-Z)P_{Z^H}^{\perp}\|_p}{\sigma_{\max}(W)}, \frac{\|(W-Z)P_{W^H}^{\perp}\|_p}{\sigma_{\max}(Z)}\right\}$$

$$\leqslant d_p(Z, W)$$

$$\leqslant \min\left\{\frac{\|(W-Z)P_{Z^H}^{\perp}\|_p}{\sigma_{\min}(W)}, \frac{\|(W-Z)P_{W^H}^{\perp}\|_p}{\sigma_{\min}(Z)}\right\}$$

$$\leqslant \|W-Z\|_p/\max\{\sigma_{\min}(Z), \sigma_{\min}(W)\}. \tag{4.51}$$

其中 $p=2$, F; $\sigma_{\max}(A)$ 与 $\sigma_{\min}(A)$ 分别表示矩阵 A 的最大与最小奇异值.

习题

1. 设 $\{A, B\}\in \mathrm{N}(n)$. $\Gamma=\operatorname{diag}(\gamma_1,\cdots,\gamma_n)$, $\Delta=\operatorname{diag}(\delta_1,\cdots,\delta_n)$, $|\gamma_i|^2+|\delta_i|^2=1$, $i=1,\cdots,n$. 又设 $\boldsymbol{x}=(\xi_1,\cdots,\xi_n)^T\in \mathbf{C}^n$, 满足 $A\boldsymbol{x}=\Gamma\boldsymbol{x}\neq 0$ 和 $B\boldsymbol{x}=\Delta\boldsymbol{x}\neq 0$. 试证: 如果对于 $\mathbf{G}_{1,2}$ 上的点 (α_0,β_0), $|\alpha_0|^2+|\beta_0|^2=1$, 有

$$(\gamma_i,\delta_i)\in \mathscr{D}_r(\alpha_0,\beta_0)\equiv\{(\alpha,\beta)\in \mathbf{G}_{1,2}: \rho((\alpha_0,\beta_0),(\alpha,\beta))\leqslant r\},$$
$$i=1,\cdots,n,$$

则必有 $(\alpha_{i_0},\beta_{i_0})\in \lambda(A,B)$, 使得 $(\alpha_{i_0},\beta_{i_0})\in \mathscr{D}_{r'}(\alpha_0,\beta_0)$, 其中 $r'=r/\sigma_{\min}((A,B))$, $\sigma_{\min}((A,B))$ 表示 $n\times 2n$ 矩阵 (A,B) 的最小奇异值.

2. 设 n 阶正则对 $\{A, B\}\in \mathrm{N}(n)$. 又设 $\boldsymbol{x}\in \mathbf{C}^n$, $\|\boldsymbol{x}\|_2=1$, $(\alpha,\beta)\in \mathbf{G}_{1,2}$, 满足 $\beta A\boldsymbol{x}\neq \alpha B\boldsymbol{x}$. 令

$$\eta=\frac{\Delta_2(A,B)}{\|X(\beta A\boldsymbol{x}-\alpha B\boldsymbol{x})\|_2},$$

其中 $\Delta_2(A,B)$ 表示 $\{A, B\}$ 的正规性偏离度 (见 (4.37) 式), $X\in \mathbf{C}^{n\times n}$ 是满足 (4.35 a)—(4.35 c) 的任一矩阵. 试证: 存在 $(\alpha_{i_0},\beta_{i_0})\in \lambda(A,B)$, 使得

$$\rho((\alpha,\beta),(\alpha_{i_0},\beta_{i_0})) \leqslant \frac{\eta}{g(\eta)} \|X(\beta A\boldsymbol{x} - \alpha B\boldsymbol{x})\|_2,$$

$g(\eta)$ 是 $g + g^2 + \cdots + g^n = \eta(\eta > 0)$ 的唯一正根.

§5 特征空间的扰动界限

5.1 特征空间

定义 5.1. 设 $\{A,B\} \in \mathbf{D}(n)$. 如果 l 维子空间 $\mathscr{X} \subset \mathbf{C}^n$ 满足

$$\dim(A\mathscr{X} + B\mathscr{X}) \leqslant \dim(\mathscr{X}), \tag{5.1}$$

则称 $\mathscr{X}$ 是 $\{A,B\}$ 的一个 l 维特征空间.

由定义 5.1 可得

推论 5.1. 设 $\{A, B\} \in \mathbf{D}(n)$, $\mathscr{X}$ 是 $\mathbf{C}^n$ 的一个子空间. 则 $\mathscr{X}$ 是 $\{A, B\}$ 的特征空间的必要与充分条件是 $\mathscr{X}$ 由 $\{A,B\}$ 的特征向量张成.

证明:

条件的充分性是显然的. 以下证明必要性.

设 $\mathscr{X}$ 是 $\{A, B\}$ 的一个 l 维特征空间. 因为对于 $\varphi \in [0, 2\pi)$, $A_\varphi\mathscr{X} + B_\varphi\mathscr{X}$ (A_φ 与 B_φ 见 (1.48) 式) 与 $A\mathscr{X} + B\mathscr{X}$ 是同一个空间, 所以应用定理 1.7, 无妨假定 B 为正定阵. 设 $\mathscr{X} = R(X_1), X_1 \in \mathbf{C}_l^{n\times l}$. 则 $\dim(B\mathscr{X}) = l$. 于是 $B\mathscr{X}$ 有一个 $n - l$ 维的正交补子空间 $\mathscr{X}'$. 由 $B > 0$ 可推知 $\mathscr{X} \oplus \mathscr{X}' = \mathbf{C}^n$. 设 $\mathscr{X}' = R(X_2)$, $X_2 \in \mathbf{C}_{n-l}^{n\times(n-l)}$, 则有 $X_2^H B X_1 = 0$. 据定义 5.1 以及 $\dim(B\mathscr{X}) = l$, 有 $A\mathscr{X} \subseteq B\mathscr{X}$. 因此 $X_2^H A X_1 = 0$. 于是得到

$$\begin{aligned} (X_1,X_2)^H A(X_1, X_2) &= \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \\ (X_1,X_2)^H B(X_1, X_2) &= \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}. \end{aligned} \tag{5.2}$$

其中 $(X_1,X_2) \in \mathbf{C}^{n\times n}$ 为非奇异阵,并且容易看出 $\{A_1,B_1\} \in \mathbf{D}(l)$. 根据定理 1.8, 存在非奇异阵 $Q_1 \in \mathbf{C}^{l\times l}$, 使得 $Q_1^H A_1 Q_1$ 与 $Q_1^H B_1 Q_1$

同时为对角阵,所以 X_1Q_1 的列是 $\{A,B\}$ 的 l 个特征向量,并且由它们张成了 $\mathscr{X}$. □

关于特征空间，还可以如下定义.

定义 5.2. 设 $\{A,B\}\in\mathbb{D}(n)$. 如果对于某一矩阵对 $\{A',B'\}\in\mathbb{D}(l)$, $1\leqslant l\leqslant n-1$, 存在 $X_1\in\mathbf{C}_l^{n\times l}$, 使得

$$AX_1B'=BX_1A', \tag{5.3}$$

则 $\mathscr{X}=R(X_1)$ 叫做 $\{A,B\}$ 的一个 l 维特征空间.

定理 5.1. 定义 5.1 与定义 5.2 等价.

证明:

1) 设 $\mathscr{X}$ 适合定义 5.1. 根据推论 5.1 的证明，存在非奇异阵 $X=(X_1,X_2)\in\mathbf{C}^{n\times n}$, 使得 $\mathscr{X}=R(X_1)$, 并且

$$X^HAX=\begin{pmatrix}A_1 & 0\\ 0 & A_2\end{pmatrix},\quad X^HBX=\begin{pmatrix}B_1 & 0\\ 0 & B_2\end{pmatrix}, \tag{5.4}$$

其中 $\{A_1,B_1\}\in\mathbb{D}(l)$, $\{A_2,B_2\}\in\mathbb{D}(n-l)$.

记 $X^{-H}=(Y_1,Y_2)$, 其中 $Y_1\in\mathbf{C}_l^{n\times l}$. (5.4) 式表明, $AX_1=Y_1A_1$, $BX_1=Y_1B_1$. 据定理 1.8, 存在非奇异阵 $Q_1\in\mathbf{C}^{l\times l}$, 使得 $Q_1^HA_1Q_1=\Lambda_1$ 与 $Q_1^HB_1Q_1=\Omega_1$ 为对角阵. 于是,若令

$$A'=Q_1\Lambda_1Q_1^H,\quad B'=Q_1\Omega_1Q_1^H,$$

则 $\{A',B'\}\in\mathbb{D}(l)$, 并且满足 $AX_1B'=BX_1A'$. 即 $\mathscr{X}=R(X_1)$ 适合定义 5.2.

2) 设 $\mathscr{X}$ 适合定义 5.2. 利用定理 1.7, 作变换(1.48), 使得变换后的 $B_\varphi>0$. 则由 (5.3) 式可得

$$A_\varphi X_1B'_\varphi=B_\varphi X_1A'_\varphi,$$

其中

$$A'_\varphi=A'\cos\varphi-B'\sin\varphi,\quad B'_\varphi=A'\sin\varphi+B'\cos\varphi,$$

并且 $\{A'_\varphi,B'_\varphi\}\in\mathbb{D}(l)$. 因此, 无妨假定 (5.3) 式中的 $B>0$.

由 $B>0$ 可推知 B' 必为非奇异阵. 事实上,如果 B' 不满秩,则根据定理 1.8, 存在非奇异阵 $Q'\in\mathbf{C}^{l\times l}$, 使得

$$A'=Q'\begin{pmatrix}\Lambda'_1 & 0\\ 0 & \Lambda'_2\end{pmatrix}Q'^H,\quad B'=Q'\begin{pmatrix}\Omega'_1 & 0\\ 0 & 0\end{pmatrix}Q'^H. \tag{5.5}$$

其中 $\Omega_1' \in \mathbf{C}_k^{k\times k}$，$k < l$，并且由 $\{A', B'\} \in \mathbf{D}(l)$ 可推知 $\Lambda_2' \in \mathbf{C}_{l-k}^{(l-k)\times(l-k)}$. 分解 $B = P^2$，$P > 0$，令

$$P^{-1}AP^{-1} = H, \ PX_1Q' = Y_1;$$

则由 (5.3) 和 (5.5) 可得

$$HY_1\begin{pmatrix} \Omega_1' & 0 \\ 0 & 0 \end{pmatrix} = Y_1\begin{pmatrix} \Lambda_1' & 0 \\ 0 & \Lambda_2' \end{pmatrix}. \tag{5.6}$$

(5.6) 式表明 $\operatorname{rank}(Y_1) < l$，即 $\operatorname{rank}(X_1) < l$，这与已设 $X_1 \in \mathbf{C}_l^{n\times l}$ 矛盾. 所以 B' 必满秩.

于是,已知的条件可叙述为

$$AX_1B' = BX_1A', \ B > 0, \ B' \in \mathbf{C}_l^{l\times l}, \ X_1 \in \mathbf{C}_l^{l\times l}.$$

由 $AX_1 = BX_1A'B'^{-1}$ 立即得到 $A\mathscr{X} \subset B\mathscr{X}$. 再由

$$\dim(B\mathscr{X}) = l = \dim(\mathscr{X})$$

得到

$$\dim(A\mathscr{X} + B\mathscr{X}) \leqslant \dim(B\mathscr{X}) = \dim(\mathscr{X}).$$

即 $\mathscr{X} = R(X_1)$ 适合定义 5.1. □

设 $\mathscr{X}$ 是 $\{A,B\} \in \mathbf{D}(n)$ 的一个 l 维特征空间. 根据特征空间的定义和上面的讨论,使得我们在讨论特征空间的扰动性质时,可以对 A, B 采用下述分解式:

$$X^HAX = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad X^HBX = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}, \tag{5.7}$$

其中

$$\left.\begin{array}{l} X = (X_1, X_2), \ X_1^HX_1 = I^{(l)}, \ X_2^HX_2 = I^{(n-l)}, \ 1 \leqslant l \leqslant n-1, \\ \qquad \{A_1, B_1\} \in \mathbf{D}(l), \ \{A_2, B_2\} \in \mathbf{D}(n-l). \end{array}\right\} \tag{5.8}$$

对于扰动后得到的 $\{\tilde{A}, \tilde{B}\} \in \mathbf{D}(n)$，我们将采用与 (5.7), (5.8) 同样的分解式,所不同者,只是在相应的符号上,加一波浪号“～”.

5.2 sinθ 第一定理

本节论证定型对的 $\sin\theta$ 第一定理,它们是 Davis-Kahan $\sin\theta$

第一定理(见第三章§7定理7.7)的推广.

定理 5.2 (定型对的 $\sin\theta$ 第一定理)[16]. 设 $\{A, B\}$ 与 $\{\widetilde{A},\widetilde{B}\}=\{A+E,B+F\}\in \mathbf{D}(n)$, 它们有 (5.7) 和 (5.8) 所示的分解式. 令 $\mathscr{X}=R(X_1)$, $\widetilde{\mathscr{X}}=R(\widetilde{X}_1)$. 如果存在满足 $\alpha+\delta\leqslant 1$ 的 $\alpha\geqslant 0$ 和 $\delta>0$, 以及 $\gamma\in\mathbf{R}$, 使得

$$\rho((\gamma,1),(\alpha_i,\beta_i))\leqslant\alpha,\ \forall(\alpha_i,\beta_i)\in\lambda(A_1,B_1) \tag{5.9}$$

和

$$\rho((\gamma,1),(\tilde{\alpha}_j,\tilde{\beta}_j))\geqslant\alpha+\delta,\ \forall(\tilde{\alpha}_j,\tilde{\beta}_j)\in\lambda(\widetilde{A}_2,\widetilde{B}_2) \tag{5.10}$$

(或者

$$\rho((\gamma,1),(\alpha_i,\beta_i)\geqslant\alpha+\delta,\ \forall(\alpha_i,\beta_i)\in\lambda(A_1,B_1)$$

和

$$\rho((\gamma,1),(\tilde{\alpha}_j,\tilde{\beta}_j))\leqslant\alpha,\forall(\tilde{\alpha}_j,\tilde{\beta}_j)\in\lambda(\widetilde{A}_2,\widetilde{B}_2)),$$

则对于任一酉不变范数$\|\ \|$,有

$$\|\sin\Theta(\mathscr{X},\widetilde{\mathscr{X}})\| \leqslant\frac{p(\alpha,\delta;\gamma)\|(A,B)\|_2}{c(A,B)\,c(\widetilde{A},\widetilde{B})}\cdot\frac{\|(EX_1,FX_1)\|}{\delta}. \tag{5.11}$$

其中

$$\Theta(\mathscr{X},\widetilde{\mathscr{X}})\equiv\arccos(X_1^H\widetilde{X}_1\widetilde{X}_1^HX_1)^{\frac{1}{2}}\geqslant 0,$$

$$p(\alpha,\delta;\gamma)=\frac{q(\gamma)[(\alpha+\delta)\sqrt{1-\alpha^2}+\alpha\sqrt{1-(\alpha+\delta)^2}\,]}{2\alpha+\delta}, \tag{5.12}$$

$$q(\gamma)=\frac{1+|\gamma|}{\sqrt{1+\gamma^2}}, \tag{5.13}$$

$$\|(A,B)\|_2=\sqrt{\|A^2+B^2\|_2},$$
$$\|(EX_1,FX_1)\|=\sqrt{\|EX_1\|^2+\|FX_1\|^2}. \tag{5.14}$$

证明:

分下列 4 步进行.

第 1 步: 建立扰动方程.

首先,令

$$Y' = X^{-H} = (Y_1', Y_2'),$$

$$\widetilde{Y}' = \widetilde{X}^{-H} = (\widetilde{Y}_1', \widetilde{Y}_2'),\ Y_1' \text{ 与 } \widetilde{Y}_1' \in \mathbf{C}^{n\times l}. \tag{5.15}$$

其中X与$\widetilde{X}$如$\{A,B\}$与$\{\widetilde{A},\widetilde{B}\}$的分解式(5.7),(5.8)所示. 进而，令

$$Y = Y'\begin{pmatrix}(Y_1'^HY_1')^{-\frac{1}{2}} & 0\\ 0 & (Y_2'^HY_2')^{-\frac{1}{2}}\end{pmatrix},$$

$$\widetilde{Y} = \widetilde{Y}'\begin{pmatrix}(\widetilde{Y}_1'^H\widetilde{Y}_1')^{-\frac{1}{2}} & 0\\ 0 & (\widetilde{Y}_2'^H\widetilde{Y}_2')^{-\frac{1}{2}}\end{pmatrix} \tag{5.16}$$

和

$$(Y_j'^HY_j')^{\frac{1}{2}}(A_j, B_j) = (A_j', B_j'),$$

$$(\widetilde{Y}_j'^H\widetilde{Y}_j')^{\frac{1}{2}}(\widetilde{A}_j, \widetilde{B}_j) = (\widetilde{A}_j', \widetilde{B}_j'),\ j = 1,2. \tag{5.17}$$

于是分解式(5.7)可改写成

$$\left.\begin{aligned}(AX, BX) &= Y\left(\begin{pmatrix}A_1' & 0\\ 0 & A_2'\end{pmatrix}, \begin{pmatrix}B_1' & 0\\ 0 & B_2'\end{pmatrix}\right),\\ (\widetilde{A}\widetilde{X}, \widetilde{B}\widetilde{X}) &= \widetilde{Y}\left(\begin{pmatrix}\widetilde{A}_1' & 0\\ 0 & \widetilde{A}_2'\end{pmatrix}, \begin{pmatrix}\widetilde{B}_1' & 0\\ 0 & \widetilde{B}_2'\end{pmatrix}\right),\end{aligned}\right\} \tag{5.18}$$

其中 $X = (X_1, X_2), Y = (Y_1, Y_2), \widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2)$ 与 $\widetilde{Y} = (\widetilde{Y}_1, \widetilde{Y}_2)$ 满足

$$\left.\begin{aligned}&X_1^HX_1 = \widetilde{X}_1^H\widetilde{X}_1 = Y_1^HY_1 = \widetilde{Y}_1^H\widetilde{Y}_1 = I^{(l)}\\ &X_2^HX_2 = \widetilde{X}_2^H\widetilde{X}_2 = Y_2^HY_2 = \widetilde{Y}_2^H\widetilde{Y}_2 = I^{(n-l)}\\ &X_1^HY_2 = \widetilde{X}_1^H\widetilde{Y}_2 = 0,\ X_2^HY_1 = \widetilde{X}_2^H\widetilde{Y}_1 = 0.\end{aligned}\right\} \tag{5.19}$$

由(5.18)可知，$AX_1 = Y_1A_1'$，$BX_1 = Y_1B_1'$. 因此可定义剩余

$$R_A = \widetilde{A}X_1 - Y_1A_1',\ R_B = \widetilde{B}X_1 - Y_1B_1'. \tag{5.20}$$

显然有关系式

$$R_A = EX_1,\ R_B = FX_1. \tag{5.21}$$

利用(5.18)和(5.15)，得到

$$\widetilde{A} = \widetilde{Y}_1\widetilde{A}_1'\widetilde{Y}_1'^H + \widetilde{Y}_2\widetilde{A}_2'\widetilde{Y}_2'^H,\ \widetilde{B} = \widetilde{Y}_1\widetilde{B}_1'\widetilde{Y}_1'^H + \widetilde{Y}_2\widetilde{B}_2'\widetilde{Y}_2'^H.$$

代入(5.20)，并取共轭转置，则有

$$\left.\begin{aligned}R_A^H &= X_1^H(\widetilde{Y}_1'\widetilde{A}_1'^H\widetilde{Y}_1^H + \widetilde{Y}_2'\widetilde{A}_2'^H\widetilde{Y}_2^H) - A_1'^HY_1^H\\ R_B^H &= X_1^H(\widetilde{Y}_1'\widetilde{B}_1'^H\widetilde{Y}_1^H + \widetilde{Y}_2'\widetilde{B}_2'^H\widetilde{Y}_2^H) - B_1'^HY_1^H.\end{aligned}\right\} \tag{5.22}$$

再利用(5.19)，(5.16)，(5.15)和(5.22)可得

$$\left.\begin{aligned}R_A^H\widetilde{X}_2 &= X_1^H\widetilde{Y}_2(\widetilde{Y}_2'^H\widetilde{Y}_2')^{\frac{1}{2}}\widetilde{A}_2'^H(\widetilde{Y}_2'^H\widetilde{Y}_2')^{-\frac{1}{2}} - A_1'^H Y_1^H\widetilde{X}_2\\ R_B^H\widetilde{X}_2 &= X_1^H\widetilde{Y}_2(\widetilde{Y}_2'^H\widetilde{Y}_2')^{\frac{1}{2}}\widetilde{B}_2'^H(\widetilde{Y}_2'^H\widetilde{Y}_2')^{-\frac{1}{2}} - B_1'^H Y_1^H\widetilde{X}_2.\end{aligned}\right\} \quad (5.23)$$

令

$$\left.\begin{aligned}\hat{A}_2 &= (\widetilde{Y}_2'^H\widetilde{Y}_2')^{-\frac{1}{2}}\widetilde{A}_2'(\widetilde{Y}_2'^H\widetilde{Y}_2')^{\frac{1}{2}} = \widetilde{A}_2(\widetilde{Y}_2'^H\widetilde{Y}_2')^{\frac{1}{2}}\\ \hat{B}_2 &= (\widetilde{Y}_2'^H\widetilde{Y}_2')^{-\frac{1}{2}}\widetilde{B}_2'(\widetilde{Y}_2'^H\widetilde{Y}_2')^{\frac{1}{2}} = \widetilde{B}_2(\widetilde{Y}_2'^H\widetilde{Y}_2')^{\frac{1}{2}},\end{aligned}\right\} \quad (5.24)$$

则(5.23)可写成

$$\begin{aligned}R_A^H\widetilde{X}_2 &= X_1^H\widetilde{Y}_2\hat{A}_2^H - A_1'^H Y_1^H\widetilde{X}_2,\\ R_B^H\widetilde{X}_2 &= X_1^H\widetilde{Y}_2\hat{B}_2^H - B_1'^H Y_1^H\widetilde{X}_2.\end{aligned} \quad (5.25)$$

再令

$$Y_1^H\widetilde{X}_2 = Z,\ X_1^H\widetilde{Y}_2 = W,\ -R_A^H\widetilde{X}_2 = C,\ -R_B^H\widetilde{X}_2 = D, \quad (5.26)$$

则扰动方程(5.25)即为

$$A_1'^H Z - W\hat{A}_2^H = C, B_1'^H Z - W\hat{B}_2^H = D, \quad (5.27)$$

其中 $A_1', B_1' \in \mathbf{C}^{l\times l}$，$\hat{A}_2$，$\hat{B}_2 \in \mathbf{C}^{(n-l)\times(n-l)}$，$Z$ 与 $W \in \mathbf{C}^{l\times(n-l)}$ 为未知矩阵.

注意到 (X_1, Y_2) 与 $(\widetilde{X}_1, \widetilde{Y}_2) \in \mathbf{C}^{n\times n}$ 为酉阵（见(5.19)式），所以根据第三章引理 7.1，有

$$\|\sin\Theta(\mathscr{X}, \widetilde{\mathscr{X}})\| = \|\widetilde{Y}_2^H X_1\| = \|W\|.$$

因此，只需证明$\|W\|$满足不等式(5 11).

第 2 步：扰动方程的简化.

为了简化方程(5.27)，首先需要对其中的矩阵 A_1'，B_1'，$\hat{A}_2$ 和 $\hat{B}_2$ 给出适当的表示.

因为分解式(5.7)中的矩阵对 $\{A_i, B_i\}(i = 1,2)$ 是定型对，所以存在非奇异阵 P_i 和实对角阵 Λ_i 与 $\Omega_i(i = 1,2)$，使得

$$\begin{aligned}&\begin{pmatrix}\Lambda_1 & 0\\ 0 & \Lambda_2\end{pmatrix} = \mathrm{diag}(\alpha_k),\ \begin{pmatrix}\Omega_1 & 0\\ 0 & \Omega_2\end{pmatrix} = \mathrm{diag}(\beta_k),\\ &(A_i, B_i) = P_i^H(\Lambda_i P_i, \Omega_i P_i),\\ &\Lambda_i^2 + \Omega_i^2 = I,\quad i = 1,2.\end{aligned} \quad (5.28)$$

令

$$Q = X\begin{pmatrix}P_1^{-1} & 0\\ 0 & P_2^{-1}\end{pmatrix} = (Q_1, Q_2),\ Q_1 \in \mathbf{C}^{n\times l}, \quad (5.29)$$

Q显然满足

$$\left.\begin{array}{r}Q^H A Q = \Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix}, \quad Q^H B Q = \Omega = \begin{pmatrix} \Omega_1 & 0 \\ 0 & \Omega_2 \end{pmatrix}, \\ \Lambda^2 + \Omega^2 = I. \end{array}\right\} \quad (5.30)$$

将 $X_i = Q_i P_i$ 代入 $X_i^H X_i = I$，得到

$$P_i P_i^H = (Q_i^H Q_i)^{-1}, \quad i = 1,2.$$

再将 P_i 的唯一分解式 $P_i = H_i V_i$ ($H_i > 0$，V_i 为酉阵)代入上式，得到 $H_i = (Q_i^H Q_i)^{-\frac{1}{2}}$，$i = 1,2$. 因此

$$P_i = (Q_i^H Q_i)^{-\frac{1}{2}} V_i, \tag{5.31}$$

$$X_i = Q_i (Q_i^H Q_i)^{-\frac{1}{2}} V_i, V_i \text{ 为酉阵}, \ i = 1,2. \tag{5.32}$$

此外,根据(5.15),

$$Y'^H Y' = (X^H X)^{-1} = \begin{pmatrix} I & X_1^H X_2 \\ X_2^H X_1 & I \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} I + X_1^H X_2 (I - X_2^H X_1 X_1^H X_2)^{-1} X_2^H X_1 & * \\ * & (I - X_2^H X_1 X_1^H X_2)^{-1} \end{pmatrix},$$

于是，有

$$\begin{aligned} Y_1'^H Y_1' &= I + X_1^H X_2 (I - X_2^H X_1 X_1^H X_2)^{-1} X_2^H X_1 \\ &= (I - X_1^H X_2 X_2^H X_1)^{-1}. \end{aligned} \tag{5.33}$$

所以,从(5.17)和(5.28)得到

$$A_1' = (Y_1'^H Y_1')^{\frac{1}{2}} P_1^H \Lambda_1 P_1, \quad B_1' = (Y_1'^H Y_1')^{\frac{1}{2}} P_1^H \Omega_1 P_1, \tag{5.34}$$

其中 P_1 和 $Y_1'^H Y_1'$ 分别由(5.31)和(5.33)给出.

类似地,对于定型对 $\{\tilde{A}_i, \tilde{B}_i\}$ ($i = 1, 2$)，存在非奇异阵 $\tilde{P}_i$ 以及实对角阵 $\tilde{\Lambda}_i$ 和 $\tilde{\Omega}_i (i = 1,2)$，使得

$$\begin{pmatrix} \tilde{\Lambda}_1 & 0 \\ 0 & \tilde{\Lambda}_2 \end{pmatrix} = \mathrm{diag}(\tilde{\alpha}_k), \quad \begin{pmatrix} \tilde{\Omega}_1 & 0 \\ 0 & \tilde{\Omega}_2 \end{pmatrix} = \mathrm{diag}(\tilde{\beta}_k),$$

$$(\tilde{A}_i, \tilde{B}_i) = \tilde{P}_i^H (\tilde{\Lambda}_i \tilde{P}_i, \ \tilde{\Omega}_i \tilde{P}_i),$$

$$\tilde{\Lambda}_i^2 + \tilde{\Omega}_i^2 = I, \ i = 1,2. \tag{5.35}$$

令

$$\widetilde{Q}=\widetilde{X}\begin{pmatrix}\widetilde{P}_1^{-1} & 0\\ 0 & \widetilde{P}_2^{-1}\end{pmatrix}=(\widetilde{Q}_1,\widetilde{Q}_2),\quad \widetilde{Q}_1\in \mathbf{C}^{n\times l},$$

$\widetilde{Q}$ 显然满足

$$\widetilde{Q}^H\widetilde{A}\widetilde{Q}=\tilde{\Lambda}=\begin{pmatrix}\tilde{\Lambda}_1 & 0\\ 0 & \tilde{\Lambda}_2\end{pmatrix},$$
$$\widetilde{Q}^H\widetilde{B}\widetilde{Q}=\tilde{\Omega}=\begin{pmatrix}\tilde{\Omega}_1 & 0\\ 0 & \tilde{\Omega}_2\end{pmatrix},\ \tilde{\Lambda}^2+\tilde{\Omega}^2=I. \tag{5.36}$$

利用与上面相同的办法，可推知

$$\widetilde{P}_i=(\widetilde{Q}_i^H\widetilde{Q}_i)^{-\frac{1}{2}}\widetilde{V}_i \tag{5.37}$$

和

$$\widetilde{X}_i=\widetilde{Q}_i(\widetilde{Q}_i^H\widetilde{Q}_i)^{-\frac{1}{2}}\widetilde{V}_i,\ \widetilde{V}_i\ \text{为酉阵},\ i=1,2 \tag{5.38}$$

以及

$$\widetilde{Y}_2'^H\widetilde{Y}_2'=(I-\widetilde{X}_2^H\widetilde{X}_1\widetilde{X}_1^H\widetilde{X}_2)^{-1}. \tag{5.39}$$

所以,从 (5.24) 和 (5.35) 得到

$$\hat{A}_2=\widetilde{P}_2^H\tilde{\Lambda}_2\widetilde{P}_2(\widetilde{Y}_2'^H\widetilde{Y}_2')^{\frac{1}{2}},\ \ \hat{B}_2=\hat{P}_2^H\tilde{\Omega}_2\widetilde{P}_2(\widetilde{Y}_2'^H\widetilde{Y}_2')^{\frac{1}{2}}, \tag{5.40}$$

其中 P_2 和 $\widetilde{Y}_2'^H\widetilde{Y}_2'$ 分别由 (5.37) 和 (5.39) 给出.

将(5.34)和(5.40)代入扰动方程(5.27)，便得到化简了的方程

$$\Lambda_1Z_1-W_1\tilde{\Lambda}_2=C_1,\ \ \Omega_1Z_1-W_1\tilde{\Omega}_2=D_1, \tag{5.41}$$

其中

$$Z_1=P_1(Y_1'^HY_1')^{\frac{1}{2}}Z\widetilde{P}_2^{-1},\ W_1=P_1^{-H}W(\widetilde{Y}_2'^H\widetilde{Y}_2')^{\frac{1}{2}}\widetilde{P}_2^H \tag{5.42}$$

和

$$C_1=P_1^{-H}C\widetilde{P}_2^{-1},\quad D_1=P_1^{-H}D\widetilde{P}_2^{-1}. \tag{5.43}$$

第 3 步：对于 $\gamma=0$ 证明不等式 (5.11).

由条件 (5.9) 和 $\alpha<1$ 可知 $\beta_i\neq 0$ 和

$$\frac{|\alpha_i/\beta_i|}{\sqrt{1-(\alpha_i/\beta_i)^2}}\leqslant\alpha,\quad 1\leqslant i\leqslant l, \tag{5.44}$$

从而

$$\left(\frac{\alpha_i}{\beta_i}\right)^2\leqslant\frac{\alpha^2}{1-\alpha^2},\quad 1\leqslant i\leqslant l.$$

再联系到 $\alpha_i^2+\beta_i^2=1$ (见 (5.28))，得到

$$|\alpha_i| \leqslant \alpha, \quad \frac{1}{|\beta_i|} \leqslant \frac{1}{\sqrt{1-\alpha^2}}, \quad 1 \leqslant i \leqslant l,$$

即

$$\|\Lambda_1\|_2 \leqslant \alpha, \quad \|\Omega_1^{-1}\|_2 \leqslant \frac{1}{\sqrt{1-\alpha^2}}. \tag{5.45}$$

类似地，由条件(5.10)和 $\alpha+\delta>0$ 可知 $\tilde{\alpha}_j \neq 0$ 和

$$\frac{1}{\sqrt{1+(\tilde{\beta}_j/\tilde{\alpha}_j)^2}} \geqslant \alpha+\delta, \quad l+1 \leqslant j \leqslant n, \tag{5.46}$$

从而

$$\left(\frac{\tilde{\beta}_j}{\tilde{\alpha}_j}\right)^2 \leqslant \frac{1-(\alpha+\delta)^2}{(\alpha+\delta)^2}, \; l+1 \leqslant j \leqslant n.$$

再联系到 $\tilde{\alpha}_j^2+\tilde{\beta}_j^2=1$ （见(5.35)），得到

$$|\tilde{\beta}_j| \leqslant \sqrt{1-(\alpha+\delta)^2},$$

$$\frac{1}{|\tilde{\alpha}_j|} \leqslant \frac{1}{\alpha+\delta}, \quad l+1 \leqslant j \leqslant n,$$

即

$$\begin{aligned} \|\tilde{\Lambda}_2^{-1}\|_2 &\leqslant \frac{1}{\alpha+\delta}, \\ \|\tilde{\Omega}_2\|_2 &\leqslant \sqrt{1-(\alpha+\delta)^2}. \end{aligned} \tag{5.47}$$

把估计式(5.45)和(5.47)与方程(5.41)联系起来，并考虑到酉不变范数必与谱范数$\| \quad \|_2$ 相容（见第二章定理 3.4)，则从

$$\|W_1\| \leqslant \|Y_1\tilde{\Lambda}_2\| \|\tilde{\Lambda}_2^{-1}\|_2 \leqslant \frac{1}{\alpha+\delta} \|W_1\tilde{\Lambda}_2\|$$

和

$$\|Z_1\| \leqslant \|\Omega_1^{-1}\|_2 \|\Omega_1 Z_1\| \leqslant \frac{1}{\sqrt{1-\alpha^2}} \|\Omega_1 Z_1\|$$

得出

$$\|C_1\| \geqslant \|W_1\tilde{\Lambda}_2\| - \|\Lambda_1 Z_1\| \geqslant (\alpha+\delta)\|W_1\| - \alpha\|Z_1\|$$

和

$$\|D_1\| \geqslant \|\Omega_1 Z_1\| - \|W_1\tilde{\Omega}_2\|$$

$$\geqslant \sqrt{1-\alpha^2}\|Z_1\| - \sqrt{1-(\alpha+\delta)^2}\|W_1\|,$$

从而

$$\|W_1\| \leqslant \frac{\|C_1\| + \alpha\|Z_1\|}{\alpha+\delta}$$

$$\leqslant \frac{\|C_1\| + \alpha(\|D_1\| + \sqrt{1-(\alpha+\delta)^2}\|W_1\|)/\sqrt{1-\alpha^2}}{\alpha+\delta}.$$

由此可得

$$\|W_1\| \leqslant \frac{(\sqrt{1-\alpha^2}\|C_1\| + \alpha\|D_1\|)[(\alpha+\delta)\sqrt{1-\alpha^2} + \alpha\sqrt{1-(\alpha+\delta)^2}]}{(\alpha+\delta)^2-\alpha^2}$$

$$\leqslant p(\alpha,\delta;0)\sqrt{\|C_1\|^2+\|D_1\|^2}/\delta. \tag{5.48}$$

再利用 $(\tilde{Y}_2'^H\tilde{Y}_2')^{-1} \leqslant I^{(n-l)}$ (见 (5.39)), 从(5.42),(5.43),(5.26)和(5.21)可导出

$$\|W\| \leqslant \|P_1\|_2\|\tilde{P}_2^{-1}\|_2\|W_1\| \tag{5.49}$$

和

$$\sqrt{\|C_1\|^2+\|D_1\|^2} \leqslant \|P_1^{-1}\|_2\|\tilde{P}_2^{-1}\|_2\sqrt{\|C\|^2+\|D\|^2}$$

$$\leqslant \|P_1^{-1}\|_2\|\tilde{P}_2^{-1}\|_2\|(EX_1, FX_1)\|. \tag{5.50}$$

把 (5.48) 和 (5.50) 代入 (5.49), 得到

$$\|W\| \leqslant p(\alpha,\delta;0)\|P_1\|_2\|P_1^{-1}\|_2\|\tilde{P}_2^{-1}\|_2^2\|(EX_1, FX_1)\|/\delta. \tag{5.51}$$

下面估计 $\|P_1\|_2, \|P_2^{-1}\|_2$ 和 $\|\tilde{P}_2^{-1}\|_2^2$.

① 从 (5.31) 和 (5.37) 可知

$$\left.\begin{aligned} &\|P_1\|_2 \leqslant \|(Q_1^HQ_1)^{-1}\|_2^{\frac{1}{2}}, \|P_1^{-1}\|_2 \leqslant \|(Q_1^HQ_1)\|_2^{\frac{1}{2}}, \\ &\|\tilde{P}_2^{-1}\|_2^2 \leqslant \|(\tilde{Q}_2^H\tilde{Q}_2)\|_2. \end{aligned}\right\} \tag{5.52}$$

② 据定理 1.7, 存在 $\varphi \in [0, 2\pi)$, 使得

$$B_\varphi = \sin\varphi A + \cos\varphi B > 0,$$

并且

$$\lambda_{\min}(B_\varphi) = c(A, B).$$

利用 (5.30), 记

$$\Omega_\varphi = \sin\varphi\Lambda + \cos\varphi\Omega = \begin{pmatrix} \Omega_{1\varphi} & 0 \\ 0 & \Omega_{2\varphi} \end{pmatrix},$$

其中 Ω_φ 是实对称阵，$0 < \Omega_\varphi \leqslant I$，$\Omega_{1\varphi} \in \mathbf{C}^{l\times l}$. 由此可得

$$c(A,B)Q_1^HQ_1 \leqslant Q_1^HB_\varphi Q_1 = \Omega_{1\varphi}$$

即

$$Q_1^HQ_1 \leqslant \Omega_{1\varphi}/c(A,B) \leqslant \frac{1}{c(A,B)} I^{(l)}, \tag{5.53}$$

从而

$$\| Q_1^HQ_1 \|_2^{\frac{1}{2}} \leqslant 1/\sqrt{c(A,B)}. \tag{5.54}$$

同理可得

$$\tilde{Q}_2^H\tilde{Q}_2 \leqslant \frac{1}{c(\tilde{A},\tilde{B})} I^{(n-l)},$$

$$\|\tilde{Q}_2^H\tilde{Q}_2\|_2 \leqslant \frac{1}{c(\tilde{A},\tilde{B})}. \tag{5.55}$$

③ 根据 (5.30)，有

$$(Q_1^HAQ_1)^2 + (Q_1^HBQ_1)^2 = I^{(l)}. \tag{5.56}$$

分解

$$Q_1 = V_1K_1, V_1 \in \mathbf{C}^{n\times l}, V_1^HV_1 = I^{(l)}, K_1 \in \mathbf{C}^{l\times l} \tag{5.57}$$

为非奇异阵，代入 (5.56)，并记

$$V_1^HAV_1 = M, \quad V_1^HBV_1 = N, \tag{5.58}$$

可得

$$(K_1K_1^H)^{-1} = MK_1K_1^HM + NK_1K_1^HN. \tag{5.59}$$

因此

$$\begin{aligned}\|(Q_1^HQ_1)^{-1}\|_2 &= \|(K_1^HK_1)^{-1}\|_2 = \|(K_1K_1^H)^{-1}\|_2 \\ &\leqslant \|M^2 + N^2\|_2\|K_1K_1^H\|_2.\end{aligned} \tag{5.60}$$

据 (5.57) 和 (5.54)，

$$\|K_1K_1^H\|_2 = \|K_1^HK_1\|_2 = \|Q_1^HQ_1\|_2 \leqslant \frac{1}{c(A,B)}, \tag{5.61}$$

据 (5.57) 和 (5.58)，

$$\|M^2 + N^2\|_2 \leqslant \|A^2 + B^2\|_2. \tag{5.62}$$

将(5.61) 和 (5.62) 代入 (5.60) 的右端，便得到

$$\|(Q_1^HQ_1)^{-1}\|_2 \leqslant \|A^2 + B^2\|_2/c(A,B), \tag{5.63}$$

④ 把 (5.54), (5.55) 和 (5.63) 代入 (5.52), 得出

$$\|P_1\|_2 \leqslant \frac{\|(A,B)\|_2}{\sqrt{c(A,B)}}, \quad \|P_1^{-1}\|_2 \leqslant \frac{1}{\sqrt{c(A,B)}},$$
$$\|\tilde{P}_2^{-1}\|_2^2 \leqslant \frac{1}{c(\tilde{A},\tilde{B})}. \tag{5.64}$$

最后把 (5.64) 代入 (5.51), 便证明了 $\gamma = 0$ 时的不等式 (5.11).

第 4 步: 对于 $\gamma \neq 0$ 证明不等式(5.11).

令

$$c = 1/\sqrt{1+\gamma^2}, \quad s = \gamma/\sqrt{1+\gamma^2},$$
$$\Lambda_{10} = \operatorname{diag}(\alpha_{i0}) = c\Lambda_1 - s\Omega_1,$$
$$\Omega_{10} = \operatorname{diag}(\beta_{i0}) = s\Lambda_1 + c\Omega_1$$

和

$$\tilde{\Lambda}_{20} = \operatorname{diag}(\tilde{\alpha}_{j0}) = c\tilde{\Lambda}_2 - s\tilde{\Omega}_2,$$
$$\tilde{\Omega}_{20} = \operatorname{diag}(\tilde{\beta}_{j0}) = s\tilde{\Lambda}_2 + c\tilde{\Omega}_2.$$

于是有

$$\alpha_{i0} = \frac{\alpha_i - \gamma\beta_i}{\sqrt{1+\gamma^2}}, \quad \beta_{i0} = \frac{\gamma\alpha_i + \beta_i}{\sqrt{1+\gamma^2}}, \quad 1 \leqslant i \leqslant l$$

和

$$\tilde{\alpha}_{j0} = \frac{\tilde{\alpha}_j - \gamma\tilde{\beta}_j}{\sqrt{1+\gamma^2}}, \quad \tilde{\beta}_{j0} = \frac{\gamma\tilde{\alpha}_j + \tilde{\beta}_j}{\sqrt{1+\gamma^2}}, \quad l+1 \leqslant j \leqslant n,$$

它们满足

$$\rho((0,1),(\alpha_{i0},\beta_{i0})) \leqslant \alpha, \quad 1 \leqslant i \leqslant l$$

和

$$\rho((0,1),(\tilde{\alpha}_{j0},\tilde{\beta}_{j0})) \geqslant \alpha + \delta, \; l+1 \leqslant j \leqslant n.$$

与此同时, 方程 (5.41) 变成了

$$\Lambda_{10}Z_1 - W_1\tilde{\Lambda}_{20} = C_{10}, \quad \Omega_{10}Z_1 - W_1\tilde{\Omega}_{20} = D_{10},$$

其中

$$C_{10} = cC_1 - sD_1, \; D_{10} = sC_1 + cD_1.$$

根据证明的第 3 步(见 (5.48) 和 (5.49)), 对任一酉不变范数$\|\ \|$,

有估计式

$$\|W\| \leqslant p(\alpha,\delta;0)\|P_1\|_2\|\tilde{P}_2^{-1}\|_2(\sqrt{1-\alpha^2}\|C_{10}\| + \alpha\|D_{10}\|)/\delta. \tag{5.65}$$

上式右端中的

$$\begin{aligned}
&\sqrt{1-\alpha^2}\|C_{10}\| + \alpha\|D_{10}\| \\
&\leqslant \sqrt{1-\alpha^2}(c\|C_1\| + |s|\|D_1\|) \\
&\quad + \alpha(|s|\|C_1\| + c\|D_1\|) \\
&= (c\sqrt{1-\alpha^2} + \alpha|s|)\|C_1\| \\
&\quad + (|s|\sqrt{1-\alpha^2} + \alpha c)\|D_1\| \\
&\leqslant [(c\sqrt{1-\alpha^2} + \alpha|s|)^2 + (|s|\sqrt{1-\alpha^2} + \alpha c)^2]^{\frac{1}{2}} \\
&\quad \times (\|C_1\|^2 + \|D_1\|^2)^{\frac{1}{2}},
\end{aligned} \tag{5.66}$$

而

$$\begin{aligned}
&(c\sqrt{1-\alpha^2} + \alpha|s|)^2 + (|s|\sqrt{1-\alpha^2} + \alpha c)^2 \\
&\quad = 1 + 4\alpha\sqrt{1-\alpha^2}\,c|s| \\
&\quad \leqslant 1 + 2c|s| \\
&\quad = 1 + \frac{2|\gamma|}{1+\gamma^2} = q^2(\gamma),
\end{aligned}$$

代入 (5.66)，再与 (5.65) 联系起来，便得到 $\gamma \neq 0$ 时的不等式 (5.11)．定理证毕．

5.3 sinθ 第二定理

对于 Frobenius 范数$\|\ \ \|_F$，不等式(5.66)可以换成不等式

$$\begin{aligned}
&\sqrt{1-\alpha^2}\|C_{10}\|_F + \alpha\|D_{10}\|_F \\
&\qquad \leqslant \sqrt{\|C_{10}\|_F^2 + \|D_{10}\|_F^2} = \sqrt{\|C_1\|_F^2 + \|D_1\|_F^2}\ ;
\end{aligned}$$

因此，(5.13)式所示的 $q(\gamma)$，可以换成 $q(\gamma) \equiv 1$．于是从(5.11)可得估计式

$$\|\sin\Theta(\mathscr{X},\tilde{\mathscr{X}})\|_F \leqslant \frac{\|(A,B)\|_2}{c(A,B)c(\tilde{A},\tilde{B})} \cdot \frac{\|(EX_1,FX_1)\|_F}{\delta}, \tag{5.67}$$

其中

$$\|(EX_1, FX_1)\|_F = \sqrt{\|EX_1\|_F^2 + \|FX_1\|_F^2}.$$

下面的定理指出：在较弱的条件下，不等式（5.67）依然成立。

定理 5.3（定型对的 $\sin\theta$ 第二定理）[16]．设 $\{A, B\}, \{\tilde{A}, \tilde{B}\}$, X 与 $\tilde{X}$ 均如定理 5.2 所述．令

$$\delta \equiv \min_{i,j}\left\{\rho((\alpha_i, \beta_i), (\tilde{\alpha}_j, \tilde{\beta}_j)):\ \begin{matrix}(\alpha_i, \beta_i) \in \lambda(A_1, B_1) \\ (\tilde{\alpha}_j, \tilde{\beta}_j) \in \lambda(\tilde{A}_2, \tilde{B}_2)\end{matrix}\right\}.$$

如果 $\delta > 0$，则不等式(5.67)成立.

证明:

如同第一章 § 5 中 5.2 所做的那样，将扰动方程(5.41)中矩阵 Z_1, W_1, C_1 和 D_1 的元素分别排列成 $l(n-l)$ 维列向量 $\boldsymbol{z}$, $\boldsymbol{w}$, $\boldsymbol{c}$ 和 $\boldsymbol{d}$, 则方程 (5.41) 可改写为(见第一章 (5.3) 式)

$$(I\otimes\Lambda_1)\boldsymbol{z} - (\tilde{\Lambda}_2\otimes I)\boldsymbol{w} = \boldsymbol{c}$$
$$(I\otimes\Omega_1)\boldsymbol{z} - (\tilde{\Omega}_2\otimes I)\boldsymbol{w} = \boldsymbol{d}.$$

由此可得

$$(\tilde{\Lambda}_2\otimes\Omega_1 - \tilde{\Omega}_2\otimes\Lambda_1)\boldsymbol{w} = -(I\otimes\Omega_1)\boldsymbol{c} + (I\otimes\Lambda_1)\boldsymbol{d}$$

和

$$\min_{\substack{1\leqslant i\leqslant l \\ l+1\leqslant j\leqslant n}} |\alpha_i\tilde{\beta}_j - \beta_i\tilde{\alpha}_j|\,\|\boldsymbol{w}\|_2 \leqslant \sqrt{\|\boldsymbol{c}\|_2^2 + \|\boldsymbol{d}\|_2^2},$$

即

$$\|W_1\|_F \leqslant \frac{\sqrt{\|C_1\|_F^2 + \|D_1\|_F^2}}{\delta}. \tag{5.68}$$

此外，根据 (5.43), (5.26) 和 (5.21)，有

$$\begin{aligned}
&\sqrt{\|C_1\|_F^2 + \|D_1\|_F^2} \\
&\qquad \leqslant \sqrt{\|P_1^{-H}R_A^H\tilde{X}_2\tilde{P}_2^{-1}\|_F^2 + \|P_1^{-H}R_B^H\tilde{X}_2\tilde{P}_2^{-1}\|_F^2} \\
&\qquad \leqslant \|P_1^{-1}\|_2\|\tilde{P}_2^{-1}\|_2\sqrt{\|R_A\|_F^2 + \|R_B\|_F^2} \\
&\qquad = \|P_1^{-1}\|_2\|\tilde{P}_2^{-1}\|_2\|(EX_1, FX_1)\|_F.
\end{aligned}$$

于是，从 (5.42) 得到

$$\|W\|_F \leqslant \|P_1\|_2\|P_1^{-1}\|_2\|\tilde{P}_2^{-1}\|_2^2\|(EX_1, FX_1)\|_F/\delta.$$

再将(5.64)代入上式右端，便导出估计式(5.67)。□

注 5.1. 由定理 5.2 和定理 5.3，可以分别推导出 Davis-Kahan $\sin\theta$ 第一定理和 $\sin\theta$ 第二定理(即第三章定理 7.7 和定理 7.9)。现以 Davis-Kahan $\sin\theta$ 第一定理为例，推导如下。

设 $A, \widetilde{A} = A + E \in \mathbf{C}^{n\times n}$ 皆为 Hermite 阵，$X = (X_1, X_2)$ 与 $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2)$ 为 n 阶酉阵，X_1 与 $\widetilde{X}_1 \in \mathbf{C}^{n\times l}(1 \leqslant l \leqslant n-1)$，使得

$$X^H A X = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad \widetilde{X}^H \widetilde{A} \widetilde{X} = \begin{pmatrix} \widetilde{A}_1 & 0 \\ 0 & \widetilde{A}_2 \end{pmatrix}, \tag{5.69}$$

其中 $A_1, \widetilde{A}_1 \in \mathbf{C}^{l\times l}$. 令 $\mathscr{X} = R(X_1), \widetilde{\mathscr{X}} = R(\widetilde{X}_1)$ 和

$$R = \widetilde{A} X_1 - X_1 A_1 \equiv E X_1 \tag{5.70}$$

以及

$$\Delta = [\alpha, \beta] \subset \mathbf{R}, \quad \Delta' = \mathbf{R} \backslash (\alpha - \delta, \beta + \delta). \tag{5.71}$$

已知

$$\lambda(A_1) \subset \Delta, \quad \lambda(\widetilde{A}_2) \subset \Delta', \tag{5.72}$$

试求 $\|\sin\Theta(\mathscr{X}, \widetilde{\mathscr{X}})\|$ 的上界.

因为同时用 $A - \frac{\alpha+\beta}{2} I$ 与 $\widetilde{A} - \frac{\alpha+\beta}{2} I$ 分别代替 A 与 $\widetilde{A}$，并不影响 R，$\mathscr{X}$ 和 $\widetilde{\mathscr{X}}$，所以无妨假定 $0 \leqslant \alpha = -\beta$.

对于 $t > 0$，考虑 $\{A, tI\}$ 与 $\{\widetilde{A}, tI\} \in \mathrm{D}(n)$. 显然有分解式

$$X^H A X = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad X^H (tI) X = \begin{pmatrix} tI & 0 \\ 0 & tI \end{pmatrix},$$

$$\widetilde{X}^H \widetilde{A} \widetilde{X} = \begin{pmatrix} \widetilde{A}_1 & 0 \\ 0 & \widetilde{A}_2 \end{pmatrix}, \quad \widetilde{X}^H (tI) \widetilde{X} = \begin{pmatrix} tI & 0 \\ 0 & tI \end{pmatrix};$$

并且，如果 $\lambda(A_1) = \{\alpha_i\}_{i=1}^l$ 和 $\lambda(\widetilde{A}_2) = \{\tilde{\alpha}_j\}_{j=l+1}^n$，则 $\lambda(A_1, tI) = \{(\alpha_i, t)\}_{i=1}^l$ 和 $\lambda(\widetilde{A}_2, tI) = \{(\tilde{\alpha}_j, t)\}_{j=l+1}^n$. 据条件(5.72)，有

$$\rho((0,1),(\alpha_i,t)) = \frac{\left|\frac{\alpha_i}{t}\right|}{\sqrt{1+\left(\frac{\alpha_i}{t}\right)^2}} \leqslant \frac{\frac{\alpha}{t}}{\sqrt{1+\left(\frac{\alpha}{t}\right)^2}}, \quad 1 \leqslant i \leqslant l$$

和

$$\rho((0,1),(\tilde{\alpha}_j,t)) = 1\bigg/\sqrt{1+\left(\frac{t}{\tilde{\alpha}_j}\right)^2}$$

$$\geqslant \frac{\alpha+\delta}{t}\bigg/\sqrt{1+\left(\frac{\alpha+\delta}{t}\right)^2},$$

$$l+1\leqslant j\leqslant n.$$

令

$$\frac{\alpha}{t}\bigg/\sqrt{1+(\alpha/t)^2}=\alpha',$$

$$\left\{\frac{\alpha+\delta}{t}\bigg/\sqrt{1+\left(\frac{\alpha+\delta}{t}\right)^2}\right\}-\left\{\frac{\alpha}{t}\bigg/\sqrt{1+\left(\frac{\alpha}{t}\right)^2}\right\}=\delta'.$$

α' 与 δ' 显然满足定理 5.2 的条件: $\alpha'\geqslant 0,\delta'>0$ 和 $\alpha'+\delta'\leqslant 1$. 因此,如果令

$$r_1(t)=\sqrt{1+\left(\frac{\alpha}{t}\right)^2},\quad r_2(t)=\sqrt{1+\left(\frac{\alpha+\delta}{t}\right)^2},$$

$$r_3(t)=\sqrt{1+\left(\frac{\|A\|_2}{t}\right)^2}$$

和

$$s_1(t)=\min_{\|x\|_2=1}\sqrt{1+\left(\frac{x^HAx}{t}\right)^2},$$

$$s_2(t)=\min_{\|x\|_2=1}\sqrt{1+\left(\frac{x^H\tilde{A}x}{t}\right)^2},$$

则从 (5.11) 稍经计算可知,对于任一酉不变范数 $\|\ \ \|$, 有不等式

$$\|\sin\Theta(\mathscr{X},\tilde{\mathscr{X}})\|$$
$$\leqslant\frac{(2\alpha+\delta)r_1(t)r_2(t)r_3(t)\|EX_1\|}{s_1(t)s_2(t)[(\alpha+\delta)r_1(t)+\alpha r_2(t)][(\alpha+\delta)r_1(t)-\alpha r_2(t)]};$$

令 $t\to+\infty$, 即得到 Davis-Kahan 的估计式(见第三章(7.27))

$$\|\sin\Theta(\mathscr{X},\tilde{\mathscr{X}})\|\leqslant\|R\|/\delta.$$

习题

1. 设 $\{A,B\}\in\mathrm{N}(n)$. 如果子空间 $\mathscr{X}\in\mathbf{C}^n$ 满足

$$\dim(A\mathscr{X} + B\mathscr{X}) \leqslant \dim(\mathscr{X}) = l,$$

则称 $\mathscr{X}$ 是 $\{A,B\}$ 的一个 l 维特征空间. 试证: $\mathbb{C}^n$ 的任一子空间 $\mathscr{X}$ 是 $\{A,B\}\in \mathrm{N}(n)$ 的特征空间的必要与充分条件是 $\mathscr{X}$ 由 $\{A,B\}$ 的特征向量张成.

2. 试由定理 5.3 导出 Davis-Kahan $\sin\theta$ 第二定理(即第三章定理 7.9).

§6 广义不变子空间的扰动界限

6.1 广义不变子空间

广义不变子空间是上节所讨论的特征空间的推广.

定义 6.1.[6] 设 $\{A,B\}$ 是 n 阶正则对, $\{A_1,B_1\}$ 是 $l(1\leqslant l\leqslant n-1)$ 阶正则对. 如果存在 $X_1\in \mathbb{C}^{n\times l}$, 适合

$$\operatorname{rank}(X_1) = l, \quad AX_1B_1 = BX_1A_1, \tag{6.1}$$

则 $\mathscr{X} = R(X_1)$ 叫做 $\{A, B\}$ 的一个 l 维广义不变子空间.

在研究与广义特征值问题 $\beta A\boldsymbol{x} = \alpha B\boldsymbol{x}$ 相联系的子空间时, G.W. Stewart[153] 引进了收缩子空间对的概念,即

定义 6.2. 设 $\{A, B\}$ 是 n 阶正则对, $\mathscr{X}$ 与 $\mathscr{Y}$ 是 $\mathbb{C}^n$ 的二子空间. 如果

$$\dim(\mathscr{X}) = \dim(\mathscr{Y}) = l$$

并且

$$A\mathscr{X}, B\mathscr{X} \subseteq \mathscr{Y},$$

则 $\mathscr{X}$ 与 $\mathscr{Y}$ 叫做 $\{A, B\}$ 的一个 l 维收缩子空间对(以下简称 l 为收缩对).

为了揭示收缩对与广义不变子空间的关系,先证明一条引理.

引理 6.1.[6] 设 $\{A,B\}$ 是 n 阶正则对. 则 $\mathscr{X}$ 是 $\{A,B\}$ 的一个 l 维广义不变子空间的必要与充分条件是存在酉阵 Y 与 $X=(X_1,X_2)\in \mathbb{C}^{n\times n}$, 其中 $X_1\in \mathbb{C}^{n\times l}(1\leqslant l\leqslant n-1)$, $\mathscr{X} = R(X_1)$, 使得

$$Y^HAX = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \quad Y^HBX = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix}, \tag{6.2}$$

其中 $A_{11}, B_{11} \in \mathbf{C}^{l\times l}$.

证明:

充分性. 由 $\{A, B\}$ 是正则对,可以立即推知 $\{A_{11}, B_{11}\}$ 是正则对. 据定理 1.1,存在满秩方阵 P_1 与 $Q_1 \in \mathbf{C}^{l\times l}$, 使得

$$A_{11} = P_1 J_{A_{11}} Q_1, \quad B_{11} = P_1 J_{B_{11}} Q_1, \tag{6.3}$$

其中 $J_{A_{11}}$ 与 $J_{B_{11}}$ 有如 (1.4) 式所示的形式. 令

$$A_1 = Q_1^{-1} J_{A_{11}}, \ B_1 = Q_1^{-1} J_{B_{11}}, \tag{6.4}$$

则由 $J_{A_{11}} J_{B_{11}} = J_{B_{11}} J_{A_{11}}$ 知 $A_{11} B_1 = B_{11} A_1$, 即

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} B_1 = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix} \begin{pmatrix} I \\ 0 \end{pmatrix} A_1. \tag{6.5}$$

将 (6.2) 代入 (6.5), 得到 $AX_1B_1 = BX_1A_1$. 据已设条件, 有 $\mathrm{rank}(X_1) = l$. 再由 (6.3) 和 (6.4) 知

$$A_1 = (P_1Q_1)^{-1} A_{11} Q_1^{-1}, \ B_1 = (P_1Q_1)^{-1} B_{11} Q_1^{-1},$$

可见 $\{A_1, B_1\}$ 是正则对. 因此, 根据定义 6.1, $\mathscr{X} = R(X_1)$ 是 $\{A, B\}$ 的一个 l 维广义不变子空间.

必要性. 首先无妨假设(6.1)式中的 X_1 适合 $X_1^H X_1 = I^{(l)}$. 根据 (6.1), 在 $\mathbf{C}^n$ 中存在一个 l 维子空间 $\mathscr{Y}$, 使得

$$A\mathscr{X}, \ B\mathscr{X} \subseteq \mathscr{Y}.$$

今在 $\mathscr{Y}$ 中取一组标准正交基, 构成 $Y_1 \in \mathbf{C}^{n\times l}$, 再在 $\mathscr{Y}^\perp$ 中取一组标准正交基,构成 $Y_2 \in \mathbf{C}^{n\times(n-l)}$, 同时在 $\mathscr{X}^\perp$ 中取一组标准正交基, 构成 $X_2 \in \mathbf{C}^{n\times(n-l)}$. 于是 $X = (X_1, X_2)$ 与 $Y = (Y_1, Y_2)$ 均为酉阵, 并且

$$Y_2^H A X_1 = Y_2^H B X_1 = 0.$$

由此立即得到分解式 (6.2). □

定理 6.1[6]. 设 $\{A, B\}$ 是 n 阶正则对, $\mathscr{X}$ 是 $\mathbf{C}^n$ 的一个 l 维子空间 $(1 \leqslant l \leqslant n-1)$. 则 $\mathscr{X}$ 是 $\{A, B\}$ 的一个广义不变子空间的必要与充分条件是存在 l 维子空间 $\mathscr{Y} \subset \mathbf{C}^n$, 使得 $\mathscr{X}$ 与 $\mathscr{Y}$ 是 $\{A, B\}$ 的收缩对.

证明:

设 $\mathscr{X}$ 与 $\mathscr{Y}$ 是 $\{A, B\}$ 的 l 维收缩对,则按照引理 6.1 中证

明必要性的办法，可证存在酉阵 $X=(X_1,X_2)$ 与 $Y=(Y_1,Y_2)$，其中 $R(X_1)=\mathscr{X}$，$R(Y_1)=\mathscr{Y}$，使得 $\{A,B\}$ 有分解式(6.2)。根据引理 6.1，$\mathscr{X}$ 是 $\{A,B\}$ 的一个广义不变子空间。

反之，如果 $\mathscr{X}$ 是 $\{A,B\}$ 的一个广义不变子空间，则由引理 6.1 知，存在酉阵 $X=(X_1,X_2)$ 与 $Y=(Y_1,Y_2)$，其中 $R(X_1)=\mathscr{X}$，使得(6.2)式成立。从而有

$$AX_1=Y_1A_{11},\quad BX_1=Y_1B_{11}. \tag{6.6}$$

记 $\mathscr{Y}=R(Y_1)$，则(6.6)式表明 $A\mathscr{X},B\mathscr{X}\subseteq\mathscr{Y}$，其中 $\dim\mathscr{Y}=l$。即 $\mathscr{X}$ 与 $\mathscr{Y}$ 构成 $\{A,B\}$ 的一个收缩对。□

6.2 算子 T(P,Q) 和函数 dif

设 $\{A,B\}$ 为 n 阶正则对，并且分块为

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},\quad B=\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix},$$

$$A_{11},B_{11}\in\mathbf{C}^{l\times l},\quad 1\leqslant l\leqslant n-1.$$

如果 $A_{21}=B_{21}=0$，则由引理 6.1 知，$R\left(\begin{pmatrix}I^{(l)}\\0\end{pmatrix}\right)$ 是 $\{A,B\}$ 的一个不变子空间。因此，如果 $(A_{21},B_{21})\neq(0,0)$，而是 $\|(A_{21},B_{21})\|$ 很小，则可以设想 $\{A,B\}$ 有一个不变子空间 $R\left(\begin{pmatrix}I^{(l)}\\P\end{pmatrix}\right)$。于是建议寻找 $P,Q\in\mathbf{C}^{(n-l)\times l}$，使得

$$\begin{pmatrix}I & 0\\ -Q & I\end{pmatrix}A\begin{pmatrix}I & 0\\ P & I\end{pmatrix}=\begin{pmatrix}* & *\\ 0 & *\end{pmatrix},$$

$$\begin{pmatrix}I & 0\\ -Q & I\end{pmatrix}B\begin{pmatrix}I & 0\\ P & I\end{pmatrix}=\begin{pmatrix}* & *\\ 0 & *\end{pmatrix}. \tag{6.7}$$

如果(6.7)式成立，则 $R\left(\begin{pmatrix}I\\P\end{pmatrix}\right)$ 显然是 $\{A,B\}$ 的一个广义不变子空间。

(6.7)式表明，P 与 Q 应满足方程组

$$\begin{cases}QA_{11}-A_{22}P=A_{21}-QA_{12}P\\ QB_{11}-B_{22}P=B_{21}-QB_{12}P.\end{cases} \tag{6.8}$$

今定义从 $\mathbf{C}^{(n-l)\times 2l}$ 到 $\mathbf{C}^{(n-l)\times 2l}$ 上的线性算子 $\mathbf{T}$:

$$\mathbf{T}(P,Q)\equiv(QA_{11}-A_{22}P,QB_{11}-B_{22}P),$$
$$P,Q\in\mathbf{C}^{(n-l)\times l},\tag{6.9}$$

同时定义算子 φ:

$$\varphi[(P,Q)]=(QA_{12}P,QB_{12}P),\tag{6.10}$$

于是，方程组 (6.8) 可写成算子方程

$$\mathbf{T}(P,Q)=(A_{21},B_{21})-\varphi[(P,Q)].\tag{6.11}$$

利用 Frobenius 范数 $\|\ \|_F$，容易看出，如果记 $\eta=\|(A_{12},B_{12})\|_F$，则第三章定理 8.1 的条件 (i) 与 (ii) 均被满足;因此,当 $\mathbf{T}$ 可逆时,可以应用第三章定理 8.1，求得方程 (6.9) 的一个解 (P,Q)，并能给出 $\|(P,Q)\|_F$ 的一个上界.

为了给出 $\|(P,Q)\|_F$ 的上界,需要用到 $\|\mathbf{T}^{-1}\|^{-1}$. 所以，必须首先研究算子 $\mathbf{T}$ 的有关性质.

定理 6.2.[153] 设 $\mathbf{T}$ 是(6.9)式定义的算子,并且假定 $\{A_{11},B_{11}\}$ 与 $\{A_{22},B_{22}\}$ 皆为正则对. 则 $\mathbf{T}$ 可逆的必要与充分条件是

$$\lambda(A_{11},B_{11})\cap\lambda(A_{22},B_{22})=\varnothing.\tag{6.12}$$

证明:

把定理的断言换一个说法,就是: 对于任意的 $R,S\in\mathbf{C}^{(n-l)\times l}$，方程组

$$QA_{11}-A_{22}P=R,\ \ QB_{11}-B_{22}P=S\tag{6.13}$$

一定有解 (P,Q) 的必要与充分条件是(6.12)式成立.

应用定理 1.1，无妨假设 A_{11}^T, B_{11}^T, A_{22} 与 B_{22} 皆有如 (1.4) 式所示的标准形式. 这时,它们都是上三角阵,仅在主对角线和上次对角线上有非零元素，并且满足

$$A_{ii}B_{ii}=B_{ii}A_{ii},\ i=1,2.$$

如同第一章 §5 中所做的那样,将 P,Q, R 与 S 的元素分别排列成 $l(n-l)$ 维列向量 $\boldsymbol{p}$, $\boldsymbol{q}$, $\boldsymbol{r}$ 与 $\boldsymbol{s}$, 则方程组 (6.13) 可改写为

$$\begin{cases}(A_{11}^T\otimes I^{(n-l)})\boldsymbol{q}-(I^{(l)}\otimes A_{22})\boldsymbol{p}=\boldsymbol{r}\\(B_{11}^T\otimes I^{(n-l)})\boldsymbol{q}-(I^{(l)}\otimes B_{22})\boldsymbol{p}=\boldsymbol{s},\end{cases}$$

即

$$T\begin{pmatrix} q \\ p \end{pmatrix}=\begin{pmatrix} r \\ s \end{pmatrix}, \tag{6.14}$$

其中

$$T=\begin{pmatrix} A_{11}^T\otimes I^{(n-l)} & -I^{(l)}\otimes A_{22} \\ B_{11}^T\otimes I^{(n-l)} & -I^{(l)}\otimes B_{22} \end{pmatrix}. \tag{6.15}$$

根据正则对的定义，存在 $\lambda\in\mathbf{C}$，使得 $\det(A_{11}+\lambda B_{11})\neq 0$。于是可作变换

$$\begin{aligned}&\begin{pmatrix} I^{(l)}\otimes I^{(n-l)} & 0 \\ -(B_{11}^T\otimes I)[(A_{11}^T+\lambda B_{11}^T)\otimes I]^{-1} & I^{(l)}\otimes I^{(n-l)} \end{pmatrix}\\ &\quad\times\begin{pmatrix} I^{(l)}\otimes I^{(n-l)} & \lambda I^{(l)}\otimes I^{(n-l)} \\ 0 & I^{(l)}\otimes I^{(n-l)} \end{pmatrix}T\\ &\quad=\begin{pmatrix} (A_{11}^T+\lambda B_{11}^T)\otimes I^{(n-l)} & * \\ 0 & T_{22} \end{pmatrix},\end{aligned} \tag{6.16}$$

其中

$$\begin{aligned}T_{22}&=(B_{11}^T\otimes I)[(A_{11}^T+\lambda B_{11}^T)\otimes I]^{-1}[I\otimes(A_{22}+\lambda B_{22})]\\ &\quad-I\otimes B_{22}\\ &=B_{11}^T(A_{11}^T+\lambda B_{11}^T)^{-1}\otimes(A_{22}+\lambda B_{22})-I\otimes B_{22}\\ &=[(A_{11}^T+\lambda B_{11}^T)^{-1}\otimes I][B_{11}^T\otimes(A_{22}+\lambda B_{22})\\ &\quad-(A_{11}^T+\lambda B_{11}^T)\otimes B_{22}]\\ &=[(A_{11}^T+\lambda B_{11}^T)^{-1}\otimes I](B_{11}^T\otimes A_{22}-A_{11}^T\otimes B_{22}).\end{aligned} \tag{6.17}$$

由(6.16)和(6.17)可知，方程组(6.14)有解 $(q^T,p^T)^T$ 的必要与充分条件是

$$\det T\equiv\det(B_{11}^T\otimes A_{22}-A_{11}^T\otimes B_{22})\neq 0. \tag{6.18}$$

设 A_{11}^T,B_{11}^T,A_{22} 与 B_{22} 的对角线元素分别为 $\{\alpha_i^{(1)}\}_{i=1}^l$，$\{\beta_i^{(1)}\}_{i=1}^l$，$\{\alpha_j^{(2)}\}_{j=1}^{n-l}$ 与 $\{\beta_j^{(2)}\}_{j=1}^{n-l}$，则显然有

$$\lambda(A_{11},B_{11})=\{(\alpha_i^{(1)},\beta_i^{(1)})\}_{i=1}^l,$$
$$\lambda(A_{22},B_{22})=\{(\alpha_j^{(2)},\beta_j^{(2)})\}_{j=1}^{n-l};$$

并且 (6.18) 式表示

$$\beta_i^{(1)}\alpha_j^{(2)}\neq\alpha_i^{(1)}\beta_j^{(2)},\quad\forall i=1,\cdots,l;\ j=1,\cdots,n-l.$$

即 $\lambda(A_{11},B_{11})\cap\lambda(A_{22},B_{22})=\varnothing$。□

如同在讨论不变子空间界限时定义函数 sep 那样(见第三章定义 8.3 与定义 8.4),可以定义正则对 $\{A_{11}, B_{11}\}$ 与 $\{A_{22}, B_{22}\}$ 的函数 dif[153]、[155]:

$$\mathrm{dif}(A_{11}, B_{11}; A_{22}, B_{22}) = \inf_{\|(P,Q)\|_F=1} \|\mathbf{T}(P,Q)\|_F. \tag{6.19}$$

其中 $\mathbf{T}(P,Q)$ 如(6.9)式所示. 容易证明,当 $\mathbf{T}$ 可逆时,有

$$\mathrm{dif}(A_{11}, B_{11}; A_{22}, B_{22}) = \|\mathbf{T}^{-1}\|^{-1}, \tag{6.20}$$

此处

$$\|\mathbf{T}^{-1}\| = \sup_{\|(P,Q)\|_F=1} \|\mathbf{T}^{-1}(P,Q)\|_F. \tag{6.21}$$

值得指出的是, $\mathrm{dif}(A_{11}, B_{11}; A_{22}, B_{22})$ 并不是集合 $\lambda(A_{11}, B_{11})$ 与 $\lambda(A_{22}, B_{22})$ 之间的距离的下界. 事实上, 对任一实数 $\sigma \neq 0$, 有 $\lambda(\sigma A_{ii}, \sigma B_{ii}) = \lambda(A_{ii}, B_{ii})$, $i = 1,2$; 但

$$\mathrm{dif}(\sigma A_{11}, \sigma B_{11}; \sigma A_{22}, \sigma B_{22}) = \sigma \mathrm{dif}(A_{11}, B_{11}; A_{22}, B_{22}).$$

因此,通过改变 σ, 就可以使 dif 很大或很小,而 $\lambda(A_{ii}, B_{ii})(i=1, 2)$ 并没有改变.

如果对 $\|P\|_F = 1$ 和 $\|Q\|_F = 1$ 先后计算 $\mathbf{T}(P,0)$ 和 $\mathbf{T}(0,Q)$, 则可得到

$$\mathrm{dif}(A_{11}, B_{11}; A_{22}, B_{22}) \leqslant \min\{\|(A_{11}, B_{11})\|_F, \|(A_{22}, B_{22})\|_F\}. \tag{6.22}$$

可见, dif 是 $\|(A_{ii}, B_{ii})\|_F$ $(i = 1,2)$ 的下界.

关于函数 dif 的性质,列举如下几条. 为了简化符号的记法,将 A_{ii} 与 B_{ii} 分别改记为 A_i 与 $B_i (i = 1,2)$.

首先一条性质是: 函数 dif 对于 A_i 与 B_i $(i = 1, 2)$ 的扰动是不敏感的,即下述定理成立.

定理 6.3.[155] 函数 dif 满足不等式

$$|\mathrm{dif}(A_1 + E_1, B_1 + F_1; A_2 + E_2, B_2 + F_2) - \mathrm{dif}(A_1, B_1; A_2, B_2)|$$
$$\leqslant \|(E_1, F_1)\|_F + \|(E_2, F_2)\|_F. \tag{6.23}$$

证明:

假定定义式 (6.19) 中的下确界对于 $\mathrm{dif}(A_1 + E_1, B_1 + F_1; A_2 + E_2, B_2 + F_2)$ 在 (P, Q) 达到,则有

$\mathrm{dif}(A_1 + E_1, B_1 + F_1; A_2 + E_2, B_2 + F_2)$

$$= \|(Q(A_1 + E_1) - (A_2 + E_2)P, Q(B_1 + F_1) - (B_2 + F_2)P)\|_F$$
$$= \|(QA_1 - A_2P, QB_1 - B_2P) + (QE_1 - E_2P, QF_1 - F_2P)\|_F$$
$$\geqslant \|(QA_1 - A_2P, QB_1 - B_2P)\|_F - \|Q(E_1, F_1) - (E_2, F_2)P\|_F$$
$$\geqslant \mathrm{dif}(A_1, B_1; A_2, B_2) - \|(E_1, F_1)\|_F - \|(E_2, F_2)\|_F.$$

同理可证

$$\mathrm{dif}(A_1, B_1; A_2, B_2) \geqslant \mathrm{dif}(A_1 + E_1, B_1 + F_1; A_2 + E_2, B_2 + F_2) - \|(E_1, F_1)\|_F - \|(E_2, F_2)\|_F.$$

所以不等式 (6.23) 成立. □

定理 6.4.[155] 如果 U_1, V_1, U_2 与 V_2 为非奇异阵,则

$$\mathrm{dif}(V_1^{-1}A_1U_1, V_1^{-1}B_1U_1; V_2^{-1}A_2U_2, V_2^{-1}B_2U_2)$$
$$\geqslant \frac{\mathrm{dif}(A_1, B_1; A_2, B_2)}{\|U_1^{-1}\|_2 \max\{\|U_1\|_2, \|V_1\|_2\} \|V_2\|_2 \max\{\|U_2^{-1}\|_2, \|V_2^{-1}\|_2\}}. \quad (6.24)$$

证明:

1) 证明

$$\mathrm{dif}(V_1^{-1}A_1U_1, V_1^{-1}B_1U_1; A_2, B_2) \geqslant \frac{\mathrm{dif}(A_1, B_1; A_2, B_2)}{\|U_1^{-1}\|_2 \max\{\|U_1\|_2, \|V_1\|_2\}}. \quad (6.25)$$

令

$$R = QV_1^{-1}A_1U_1 - A_2P, S = QV_1^{-1}B_1U_1 - B_2P.$$

取 (P, Q) 满足

(i) $\|(P, Q)\|_F = 1$,

(ii) $\|(R, S)\|_F = \mathrm{dif}(V_1^{-1}AU_1, V_1^{-1}B_1U_1; A_2, B_2)$.

于是有

$$\|(R, S)\|_F \|U_1^{-1}\|_2 \geqslant \|(RU_1^{-1}, SU_1^{-1})\|_F$$
$$= \|(QV_1^{-1}A_1 - A_2PU_1^{-1}, QV_1^{-1}B_1 - B_2PU_1^{-1})\|_F$$

$$\geqslant \operatorname{dif}(A_1, B_1; A_2, B_2)\|(PU_1^{-1}, QV_1^{-1})\|_F$$
$$\geqslant \operatorname{dif}(A_1, B_1; A_2, B_2)/\max\{\|U_1\|_2, \|V_1\|_2\}.$$

因此，不等式 (6.25) 成立.

2）证明

$$\operatorname{dif}(A_1, B_1; V_2^{-1}A_2U_2, V_2^{-1}B_2U_2) \geqslant \frac{\operatorname{dif}(A_1, B_1; A_2, B_2)}{\|V_2\|_2\max\{\|U_2^{-1}\|_2, \|V_2^{-1}\|_2\}}. \quad (6.26)$$

令

$$R = QA_1 - V_2^{-1}A_2U_2P, \ S = QB_1 - V_2^{-1}B_2U_2P.$$

取 (P, Q) 满足

(i) $\|(P, Q)\|_F = 1$,

(ii) $\|(R, S)\|_F = \operatorname{dif}(A_1, B_1; V_2^{-1}A_2U_2, V_2^{-1}B_2U_2)$.

于是有

$$\begin{aligned}\|V_2\|_2\|(R, S)\|_F &\geqslant \|(V_2R, V_2S)\|_F \\ &= \|(V_2QA_1 - A_2U_2P, V_2QB_1 - B_2U_2P)\|_F \\ &\geqslant \operatorname{dif}(A_1, B_1; A_2, B_2)\cdot\|(U_2P, V_2Q)\|_F \\ &\geqslant \operatorname{dif}(A_1, B_1; A_2, B_2)/\max\{\|U_2^{-1}\|_2, \|V_2^{-1}\|_2\}.\end{aligned}$$

因此，不等式 (6.26) 成立.

利用 (6.25) 与 (6.26)，立即得到不等式 (6.24). □

定理 6.5.[155] 设

$$A_1 = \operatorname{diag}(A_1^{(1)}, \cdots, A_k^{(1)}), \quad B_1 = \operatorname{diag}(B_1^{(1)}, \cdots, B_k^{(1)}),$$
$$A_2 = \operatorname{diag}(A_1^{(2)}, \cdots, A_l^{(2)}), \quad B_2 = \operatorname{diag}(B_1^{(2)}, \cdots, B_l^{(2)}).$$

则

$$\operatorname{dif}(A_1, B_1; A_2, B_2) = \min_{\substack{1\leqslant i\leqslant k \\ 1\leqslant j\leqslant l}} \operatorname{dif}(A_i^{(1)}, B_i^{(1)}; A_j^{(2)}, B_j^{(2)}). \quad (6.27)$$

证明:

只需证明

$$\operatorname{dif}\left(\begin{pmatrix} A_1^{(1)} & 0 \\ 0 & A_2^{(1)} \end{pmatrix}, \begin{pmatrix} B_1^{(1)} & 0 \\ 0 & B_2^{(1)} \end{pmatrix}; A_2, B_2\right)$$
$$= \min\{\operatorname{dif}(A_1^{(1)}, B_1^{(1)}; A_2, B_2), \operatorname{dif}(A_2^{(1)}, B_2^{(1)}; A_2, B_2)\}; \quad (6.28)$$

因为同理可证

$$\operatorname{dif}\left(A_1,\ B_1;\ \begin{pmatrix} A_1^{(2)} & 0 \\ 0 & A_2^{(2)} \end{pmatrix},\ \begin{pmatrix} B_1^{(2)} & 0 \\ 0 & B_2^{(2)} \end{pmatrix}\right)$$

$$= \min\{\operatorname{dif}(A_1, B_1; A_1^{(2)}, B_1^{(2)}), \operatorname{dif}(A_1, B_1; A_2^{(2)}, B_2^{(2)})\},$$

然后可以归纳地证明等式(6.27).

首先指出,

$$\lambda\left(\begin{pmatrix} A_1^{(1)} & 0 \\ 0 & A_2^{(1)} \end{pmatrix},\ \begin{pmatrix} B_1^{(1)} & 0 \\ 0 & B_2^{(1)} \end{pmatrix}\right) \cap \lambda(A_2, B_2) \neq \varnothing$$

的必要与充分条件是

$$\lambda(A_1^{(1)}, B_1^{(1)}) \cap \lambda(A_2, B_2) \neq \varnothing$$

或

$$\lambda(A_2^{(1)}, B_2^{(1)}) \cap \lambda(A_2, B_2) \neq \varnothing.$$

因此,(6.28)式两端必同时为零,或同时不为零. 现假设(6.28)两端同时不为零. 这时,有

$$\operatorname{dif}^2\left(\begin{pmatrix} A_1^{(1)} & 0 \\ 0 & A_2^{(1)} \end{pmatrix},\ \begin{pmatrix} B_1^{(1)} & 0 \\ 0 & B_2^{(1)} \end{pmatrix};\ A_2,\ B_2\right)$$

$$= \inf_{\|(P,Q)\|_F=1} \left\| \left((Q_1, Q_2) \begin{pmatrix} A_1^{(1)} & 0 \\ 0 & A_2^{(1)} \end{pmatrix} - A_2(P_1, P_2),\ (Q_1, Q_2) \begin{pmatrix} B_1^{(1)} & 0 \\ 0 & B_2^{(1)} \end{pmatrix} - B_2(P_1, P_2) \right) \right\|_F^2$$

$$= \inf_{\|(P,Q)\|_F=1} \{ \|(Q_1A_1^{(1)} - A_2P_1, Q_1B_1^{(1)} - B_2P_1)\|_F^2 + \|(Q_2A_2^{(1)} - A_2P_2, Q_2B_2^{(1)} - B_2P_2)\|_F^2 \}$$

$$\geqslant \inf_{\|(P_1,Q_1)\|_F^2 + \|(P_2,Q_2)\|_F^2 = 1} \{ \operatorname{dif}^2(A_1^{(1)}, B_1^{(1)}; A_2, B_2) \|(P_1, Q_1)\|_F^2 + \operatorname{dif}^2(A_2^{(1)}, B_2^{(1)}; A_2, B_2) \|(P_2, Q_2)\|_F^2 \}$$

$$\geqslant \min\{\operatorname{dif}^2(A_1^{(1)}, B_1^{(1)}; A_2, B_2), \operatorname{dif}^2(A_2^{(1)}, B_2^{(1)}; A_2, B_2)\}.$$

另一方面, 无妨设

$$\operatorname{dif}(A_1^{(1)}, B_1^{(1)}; A_2, B_2) \leqslant \operatorname{dif}(A_2^{(1)}, B_2^{(1)}; A_2, B_2).$$

取 P_1 与 Q_1 满足 $\|(P_1, Q_1)\|_F = 1$, 并且

$$\operatorname{dif}(A_1^{(1)}, B_1^{(1)}; A_2, B_2)$$

$$= \|(Q_1A_1^{(1)} - A_2P_1, Q_1B_1^{(1)} - B_2^{(1)}P_1)\|_F.$$

则

$$\begin{aligned}
&\operatorname{dif}(A_1, B_1; A_2, B_2) \\
&\leqslant \inf_{\|(P_1,Q_1)\|_F=1} \left\| \left((Q_1, 0)\begin{pmatrix} A_1^{(1)} & 0 \\ 0 & A_2^{(1)} \end{pmatrix} - A_2(P_1, 0), \right.\right. \\
&\quad \left.\left. \times (Q_1, 0)\begin{pmatrix} B_1^{(1)} & 0 \\ 0 & B_2^{(1)} \end{pmatrix} - B_2(P_1, 0) \right) \right\|_F \\
&= \|(Q_1A_1^{(1)} - A_2P_1, Q_1B_1^{(1)} - B_2P_1)\|_F \\
&= \operatorname{dif}(A_1^{(1)}, B_1^{(1)}; A_2, B_2).
\end{aligned}$$

因此 (6.28) 式得证. □

6.3 逼近定理与扰动定理

应用第三章定理 8.1 可知，对于

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

和 (6.9) 式所示的算子 $\mathbf{T}$, 如果 $\{A, B\}$ 和 $\{A_{ii}, B_{ii}\}$ $(i = 1,2)$ 皆为正则对，$\mathbf{T}$ 可逆，并且令 $\delta = \|\mathbf{T}^{-1}\|^{-1}$, $\gamma = \|(A_{21}, B_{21})\|_F$, $\eta = \|(A_{12}, B_{12})\|_F$, 则当 $\eta\gamma/\delta^2 < 1/4$ 时，必存在满足 $\|(P, Q)\|_F < 2\gamma/\delta$ 的 (P, Q), 使得 $R\left(\begin{pmatrix} I \\ P \end{pmatrix}\right)$ 是 $\{A, B\}$ 的广义不变子空间. 把 $\|P\|_F \leqslant \|(P,Q)\|_F < 2\gamma/\delta$ 与第三章的 (8.6) 和 (8.9) 式联系起来，就刻划了 $R\left(\begin{pmatrix} I \\ 0 \end{pmatrix}\right)$ 作为 $\{A, B\}$ 的近似广义不变子空间的逼近程度. 这一结论可叙述如下:

定理 6.6 (逼近定理)[153]. 设 $\{A, B\}$ 为 n 阶正则对，$X = (X_1, X_2)$ 与 $Y = (Y_1, Y_2)$ 是酉阵，X_1 与 $Y_1 \in \mathbf{C}^{n\times l}$ $(1 \leqslant l \leqslant n-1)$. 记

$$Y^HAX = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

$$Y^HBX = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \quad A_{11}, B_{11} \in \mathbf{C}^{l\times l}.$$

其中 $\{A_{ii}, B_{ii}\}(i=1,2)$ 皆为正则对．令

$$\gamma = \|(A_{21}, B_{21})\|_F, \quad \eta = \|(A_{12}, B_{12})\|_F.$$

如果

$$\delta \equiv \text{dif}\,(A_{11}, B_{11}; A_{22}, B_{22}) > 0,$$

并且

$$\frac{\eta\gamma}{\delta^2} < \frac{1}{4},$$

则存在 $P \in \mathbf{C}^{(n-l)\times l}$，满足 $\|P\|_F < \frac{2\gamma}{\delta}$，使得 $R\left(X\begin{pmatrix} I \\ P \end{pmatrix}\right)$ 是 $\{A, B\}$ 的广义不变子空间.

定理 6.6 表明，$R\left(X\begin{pmatrix} I \\ 0 \end{pmatrix}\right)$（即 $R(X_1)$）是 $\{A,B\}$ 的一个近似广义不变子空间，并且 $\|P\|_F < \frac{2\gamma}{\delta}$ 反映了它和 $\{A,B\}$ 的广义不变子空间 $R\left(X\begin{pmatrix} I \\ P \end{pmatrix}\right)$ 的差距.

如果把 $\{A,B\}$ 的广义不变子空间看作 $\{A+E, B+F\}$ 的近似广义不变子空间，则可从定理 6.6 得到下述扰动定理.

定理 6.7（扰动定理）[153]. 设 $\{A, B\}$ 与 $\{A+E, B+F\}$ 皆为 n 阶正则对，$X=(X_1, X_2)$ 与 $Y=(Y_1, Y_2)$ 是酉阵，X_1 与 $Y_1 \in \mathbf{C}^{n\times l}(1 \leqslant l \leqslant n-1)$，满足

$$Y^H A X = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

$$Y^H B X = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix}, \quad A_{11}, B_{11} \in \mathbf{C}^{l\times l},$$

即 $\mathscr{X} = R(X_1)$ 是 $\{A, B\}$ 的一个 l 维广义不变子空间．记

$$Y^H E X = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix},$$

$$Y^H F X = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}, \quad E_{11},\ F_{11} \in \mathbf{C}^{l\times l}.$$

令

$$\gamma = \|(E_{21}, F_{21})\|_F, \quad \eta = \|(A_{12}, B_{12})\|_F + \|(E_{12}, F_{12})\|_F.$$

如果

$$\delta \equiv \mathrm{dif}(A_{11}, B_{11}; A_{22}, B_{22}) - \|(E_{11}, F_{11})\|_F - \|(E_{22}, F_{22})\|_F > 0,$$

并且

$$\frac{\eta\gamma}{\delta^2} < \frac{1}{4},$$

则存在 $P \in \mathbb{C}^{(n-l)\times l}$，满足 $\|P\|_F < \frac{2\gamma}{\delta}$，使得 $R\left(X\begin{pmatrix} I \\ P \end{pmatrix}\right)$是$\{A+E, B+F\}$ 的广义不变子空间.

下面用一个最简单的情况，说明定理 6.7 的结论.

推论 6.1. 设 $\{A, B\}$ 是 n 阶正则对，$E, F \in \mathbb{C}^{n\times n}$，$(\alpha_1, \beta_1)$ 是 $\{A, B\}$ 的单特征值，$\boldsymbol{x}$ 是相应的单位特征向量. 设 $X = (\boldsymbol{x}_1, X_2)$ 与 $Y = (\boldsymbol{y}_1, Y_2)$ 是酉阵，使得

$$Y^H AX = \begin{pmatrix} \tau\alpha_1 & \boldsymbol{a}^H \\ 0 & A_{22} \end{pmatrix}, \quad Y^H BX = \begin{pmatrix} \tau\beta_1 & \boldsymbol{b}^H \\ 0 & B_{22} \end{pmatrix}, \quad \tau \neq 0,$$

$$Y^H EX = \begin{pmatrix} e_{11} & \boldsymbol{h}^H \\ \boldsymbol{g} & E_{22} \end{pmatrix}, \quad Y^H FX = \begin{pmatrix} f_{11} & \boldsymbol{s}^H \\ \boldsymbol{r} & F_{22} \end{pmatrix}.$$

令

$$\gamma = \|(\boldsymbol{g}^T, \boldsymbol{r}^T)^T\|_2, \quad \eta = \|(\boldsymbol{a}^T, \boldsymbol{b}^T)^T\|_2 + \|(\boldsymbol{h}^H, \boldsymbol{s}^H)^T\|_2,$$

$$\varepsilon = \|(E, F)\|_2$$

和

$$\delta = \left\|\begin{pmatrix} \alpha_1 I & A_{22} \\ \beta_1 I & B_{22} \end{pmatrix}^{-1}\right\|_2^{-1}.$$

如果

$$\delta - 2\varepsilon > 0,$$

并且

$$\frac{\eta\gamma}{(\delta - 2\varepsilon)^2} < \frac{1}{4},$$

则存在 $\boldsymbol{p} \in \mathbb{C}^{n-1}$，满足 $\|\boldsymbol{p}\|_2 < \frac{2\gamma}{\delta - 2\varepsilon}$，使得

$$\tilde{\boldsymbol{x}}_1 = \boldsymbol{x}_1 + X_2\boldsymbol{p}$$

是 $\{A+E, B+F\}$ 的特征向量,并且 $R(\tilde{\boldsymbol{x}}_1)$ 与 $R(\boldsymbol{x}_1)$ 之间的距离

$$\sin\theta(R(\tilde{\boldsymbol{x}}_1), R(\boldsymbol{x}_1)) \leqslant \frac{2\gamma}{\sqrt{(\delta-2\varepsilon)^2+4\gamma^2}}.$$

证明:

直接应用定理 6.7 即可. 只需说明一点,即

$$\delta = \mathrm{dif}(\alpha_1,\beta_1;A_{22},B_{22}).$$

这是因为,据题设, $(\alpha_1,\beta_1)\bar{\in}\lambda(A_{22},B_{22})$, 所以按照定义,有

$$\mathrm{dif}(\alpha_1,\beta_1;A_{22},B_{22}) = \|\mathrm{T}^{-1}(\boldsymbol{p},\boldsymbol{q})\|^{-1},$$

其中

$$\mathrm{T}(\boldsymbol{p},\boldsymbol{q}) \equiv (\boldsymbol{q}\alpha_1 - A_{22}\boldsymbol{p}, \boldsymbol{q}\beta_1 - B_{22}\boldsymbol{p}).$$

于是有

$$\begin{aligned}
\|\mathrm{T}^{-1}(\boldsymbol{p},\boldsymbol{q})\|^{-1} &= \left[\sup_{(\boldsymbol{p},\boldsymbol{q})\neq 0}\frac{\|\mathrm{T}^{-1}(\boldsymbol{p},\boldsymbol{q})\|_F}{\|(\boldsymbol{p},\boldsymbol{q})\|_F}\right]^{-1} \\
&= \left[\sup_{(\boldsymbol{p},\boldsymbol{q})\neq 0}\frac{\|(\boldsymbol{p},\boldsymbol{q})\|_F}{\|\mathrm{T}(\boldsymbol{p},\boldsymbol{q})\|_F}\right]^{-1} \\
&= \inf_{(\boldsymbol{p},\boldsymbol{q})\neq 0}\frac{\|\mathrm{T}(\boldsymbol{p},\boldsymbol{q})\|_F}{\|(\boldsymbol{p},\boldsymbol{q})\|_F} \\
&= \inf_{\binom{\boldsymbol{q}}{\boldsymbol{p}}\neq 0}\left\{\left\|\begin{pmatrix}\alpha_1 I & -A_{22}\\ \beta_1 I & -B_{22}\end{pmatrix}\begin{pmatrix}\boldsymbol{q}\\ \boldsymbol{p}\end{pmatrix}\right\|_2 \Big/ \left\|\begin{pmatrix}\boldsymbol{q}\\ \boldsymbol{p}\end{pmatrix}\right\|_2\right\} \\
&= \left\|\begin{pmatrix}\alpha_1 I & A_{22}\\ \beta_1 I & B_{22}\end{pmatrix}^{-1}\right\|_2^{-1}. \quad \square
\end{aligned}$$

注 6.1. 关于与奇异值分解以及广义奇异值分解相联系的子空间(即所谓奇异子空间和广义奇异子空间)的研究,已经取得一些基本的结果,读者可参阅 Wedin, Stewart 和作者的有关论文[15]、[155]和[182].

习题

1. 设 $A_1 \in \mathbf{C}^{l\times l}$, $A_2 \in \mathbf{C}^{m\times m}$. 定义

$$\mathrm{sep}_F(A_1, A_2) = \inf_{\substack{P \in \mathbb{C}^{m\times l} \\ \|P\|_F=1}} \|PA_1 - A_2P\|_F.$$

试证:

(1) $\dfrac{1}{\sqrt{2}}\mathrm{sep}_F(A_1, A_2) \geqslant \mathrm{dif}(A_1, I^{(l)}; A_2, I^{(m)})$,

(2) 如果 $\|A_1\|_2 \leqslant 1$, $\|A_2\|_2 \leqslant 1$, 则

$$\mathrm{dif}(A_1, I^{(l)}; A_2, I^{(m)}) \geqslant \frac{1}{2}\mathrm{sep}_F(A_1, A_2).$$

2. 设 $\{A_1, B_1\}$ 为 l 阶正则对，$\{A_2, B_2\}$ 为 m 阶正则对. 定义算子 $\mathbf{T}$:

$$\mathbf{T}(P, Q) = (QA_1 - A_2P, QB_1 - B_2P),\ P, Q \in \mathbb{C}^{m\times l}.$$

将 (P, Q) 的元素按某一固定的顺序排列为 $2lm$ 维列向量，同时把 $\mathbf{T}$ 看作 $\mathbb{C}^{2lm}$ 内的一个线性变换. 试证: 如果 T 是 $\mathbf{T}$ 的矩阵表示，则

$$\frac{1}{\mathrm{dif}(A_1, B_1;\ A_2, B_2)} \leqslant \|T^{-1}\|_2.$$

第四章说明

关于正则矩阵对广义特征值问题扰动性质的研究，是从七十年代开始的. Stewart 在七十年代发表了论文[153]、[155]、[156]、[158]和[159]，在这些论文中，他分析了研究广义特征值问题扰动理论的必要性，并且第一个使用弦度量，建立了广义特征值的 Gerschgorin 理论，求出了定型对的广义特征值的扰动界限(本章推论 3.1)，分析了收缩子空间对的扰动规律（本章定理 6.6—6.7). 作者在此基础上，作了进一步的研究. 关于正则矩阵对广义特征值的扰动，作者的着眼点是把矩阵对放到复投影空间中去考察，先用某种投影度量去界定广义特征值的扰动，然后再把投影度量与通常的 Euclid 度量联系起来. 对于广义奇异值的扰动，可以类似地进行研究. 关于广义不变子空间的扰动，作者把 Davis-Kahan 理论推广到定型对的情形，解决了 Moler 和 Stewart 在 1977 年

提出的一个 open question[159]，同时阐明了研究广义不变子空间扰动性质的一种办法．有关这方面的问题，读者可参阅作者的论文[8]—[11]、[13]、[15]和[16]．

关于奇异对的广义特征值问题的扰动分析，目前的工作还不多．可参看作者的论文[14]、[22]和 Van Dooren 的论文[174]．

关于广义特征值问题扰动理论的研究，还存在不少问题．例如：广义特征值问题条件数的性质和估计，根据不同需要和不同类型的矩阵对，来求扰动界限，以及奇异对的广义特征值问题的扰动分析，等等，都有待于深入研究．

第五章　广义逆与最小二乘问题扰动分析

§1　矩阵逆与线性方程组解的扰动

先看一个例子.

例 1.1　$A=\begin{pmatrix}2 & 6\\ 2 & 6.00001\end{pmatrix}$,　$\widetilde{A}=\begin{pmatrix}2 & 6\\ 2 & 5.99999\end{pmatrix}$,

$$\boldsymbol{b}=\begin{pmatrix}8\\ 8.00001\end{pmatrix},\quad \tilde{\boldsymbol{b}}=\begin{pmatrix}8\\ 8.00002\end{pmatrix}.$$

计算可知

$$A^{-1}=\begin{pmatrix}300000.5 & -300000\\ -100000 & 100000\end{pmatrix},$$

$$\widetilde{A}^{-1}=\begin{pmatrix}-299999.5 & 300000\\ 100000 & -100000\end{pmatrix};$$

方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 的解 $\boldsymbol{x}=(1,1)^T$，方程组 $\widetilde{A}\tilde{\boldsymbol{x}}=\tilde{\boldsymbol{b}}$ 的解 $\tilde{\boldsymbol{x}}=(10,-2)^T$.

例 1.1 说明，A 和 $\boldsymbol{b}$ 虽然经过很小的扰动，但 A 的逆和 $A\boldsymbol{x}=\boldsymbol{b}$ 的解 $\boldsymbol{x}$ 却都产生了很大的偏差. 因此有必要研究非奇异阵的逆和线性方程组的解的稳定性.

本节将证明，非奇异阵的逆是矩阵元素的连续函数，线性方程组的解是系数矩阵和右端项的元素的连续函数；同时给出扰动之后的矩阵的逆和线性方程组的解的扰动界限.

在本节和本章下面几节，我们将使用由 $\mathbf{R}^N$ 上的 SG 函数一致生成的、在 $\bigcup_{m,n=1}^{N}\mathbf{C}^{m\times n}$ 上相容的酉不变范数$\|\ \ \|$，其中 N 是一个足够大的自然数

在第三章 §8 的 8.1 中，我们已经指出，如果$\|\ \ \|$是 $\mathbf{C}^{N\times N}$ 上的

酉不变范数，它由 R^N 上的 SG 函数 $\Phi_N(\xi_1,\cdots,\xi_N)$ 生成，则可如下定义一个在 $\bigcup_{m,n=1}^{N}\mathbf{C}^{m\times n}$ 上相容的酉不变范数$\|\ \ \|$：对于任一 $A\in\mathbf{C}^{m\times n}$ $(m,\ n\leqslant N)$，设 A 的奇异值分解为 $A=U\Sigma V^H$，其中 U 与 V 为酉阵，$\Sigma=\mathrm{diag}(\sigma_1,\sigma_2,\cdots)$，$\sigma_1\geqslant\cdots\geqslant\sigma_n$ 是 A 的奇异值，则定义

$$\|A\|=\Phi_N(\sigma_1,\cdots,\sigma_n,0,\cdots,0).$$

这样得到的范数，就叫做由 R^N 上的 SG 函数 Φ_N 一致生成的、在 $\bigcup_{m,n=1}^{N}\mathbf{C}^{m\times n}$ 上相容的酉不变范数.

这种范数有下述一些重要的性质（这些性质可以直接从范数$\|\ \ \|$的上述定义，并根据第二章 § 3 中 3.4 所述的酉不变范数的性质导出）：对于一切自然数 $m,n,\ l\leqslant N$，有

1) $$\|A\|=\left\|\begin{pmatrix}0&0&0\\0&A&0\\0&0&0\end{pmatrix}\right\|,\ \forall\begin{pmatrix}0&0&0\\0&A&0\\0&0&0\end{pmatrix}\in\mathbf{C}^{m\times n},\quad(1.1)$$

2) $$\|A^H\|=\|A\|,\ \forall A\in\mathbf{C}^{m\times n}.\quad(1.2)$$

3) $$\|A\|=\|A\|_2,\ \forall A\in\mathbf{C}^{m\times n},\ \mathrm{rank}(A)=1;\quad(1.3)$$

特别地，

$$\|\boldsymbol{x}\|=\|\boldsymbol{x}\|_2,\ \forall\boldsymbol{x}\in\mathbf{C}^n.\quad(1.4)$$

4) $$\|CD\|\leqslant\|C\|_2\|D\|,\ \|CD\|\leqslant\|C\|\|D\|_2,$$
$$\forall C\in\mathbf{C}^{m\times n},\ D\in\mathbf{C}^{n\times l}.\quad(1.5)$$

5) $$\|A\|_2\leqslant\|A\|,\forall A\in\mathbf{C}^{m\times n}.\quad(1.6)$$

不等式 (1.6) 可如下导出．由 (1.4) 和 (1.5) 可知

$$\|A\boldsymbol{x}\|_2=\|A\boldsymbol{x}\|\leqslant\|A\|\|\boldsymbol{x}\|_2,$$

再据 $\|A\|_2$ 的定义 $\|A\|_2=\max_{\|\boldsymbol{x}\|_2=1}\|A\boldsymbol{x}\|_2$，所以 (1.6) 式成立.

下述不等式是(1.5)与 (1.6) 的推论.

6) $$\|CD\|\leqslant\|C\|\|D\|.\quad(1.7)$$

此外还有

7） 设 $C\in\mathbf{C}^{m\times n}, D\in\mathbf{C}^{l\times n}$. 如果

$$\|Cx\|_2\leqslant\|Dx\|_2, \forall x\in\mathbf{C}^n,$$

则 $\|C\|\leqslant\|D\|$.

这是因为，矩阵的奇异值具有 minimax 性质（见第三章 §3 习题 2）：设 $\sigma_1(A)\geqslant\sigma_2(A)\geqslant\cdots$ 是 A 的奇异值，则

$$\sigma_i(A)=\max_{\substack{\mathscr{X}\\ \dim(\mathscr{X})=i}}\min_{\substack{x\in\mathscr{X}\\ \|x\|_2=1}}\|Ax\|_2,\quad i=1,2,\cdots.$$

由 $\|Cx\|_2\leqslant\|Dx\|_2, \forall x\in\mathbf{C}^n$，可导出 $\sigma_i(C)\leqslant\sigma_i(D), i=1,\cdots,n$. 再根据第二章定理 3.3，便得到 $\|C\|\leqslant\|D\|$.

1.1 矩阵逆的扰动界限

首先证明一个简单的结论.

定理 1.1. 如果 $A\in\mathbf{C}^{n\times n}$ 与 $B=A+E$ 都是非奇异阵，则

$$\frac{\|B^{-1}-A^{-1}\|}{\|A^{-1}\|}\leqslant\bar{\kappa}\frac{\|E\|_2}{\|A\|},\tag{1.8}$$

其中

$$\bar{\kappa}=\|A\|\|B^{-1}\|_2.\tag{1.9}$$

证明:

由

$$B^{-1}-A^{-1}=A^{-1}(A-B)B^{-1}\tag{1.10}$$

立即得到

$$\|B^{-1}-A^{-1}\|\leqslant\|A^{-1}\|\|A-B\|_2\|B^{-1}\|_2,$$

即

$$\frac{\|B^{-1}-A^{-1}\|}{\|A^{-1}\|}\leqslant\|A\|\ \|B^{-1}\|_2\frac{\|E\|_2}{\|A\|},$$

这正是不等式 (1.8). □

定理 1.2. 设 $A\in\mathbf{C}^{n\times n}$ 是非奇异阵，$E\in\mathbf{C}^{n\times n}$. 如果

$$\|A^{-1}\|_2\|E\|_2<1,\tag{1.11}$$

则 $B=A+E$ 必是非奇异阵，并且有

$$\|B^{-1}\|\leqslant\|A^{-1}\|/\gamma\tag{1.12}$$

和

$$\frac{\|B^{-1}-A^{-1}\|}{\|A^{-1}\|} \leqslant \frac{\kappa}{\gamma} \frac{\|E\|_2}{\|A\|}, \tag{1.13}$$

其中

$$\kappa = \|A\| \|A^{-1}\|_2 \tag{1.14}$$

和

$$\gamma = 1 - \kappa \|E\|_2 / \|A\| > 0. \tag{1.15}$$

证明:

1）首先注意到

$$A + E = A(I + A^{-1}E),$$

其中 $\|A^{-1}E\| \leqslant \|A^{-1}\|_2 \|E\| < 1$. 因此根据第二章定理 2.4，$I + A^{-1}E$ 是非奇异阵，从而 $B = A + E$ 是非奇异阵.

于是得到

$$\begin{aligned} \|B^{-1}\| &= \|(A+E)^{-1}\| = \|(I + A^{-1}E)^{-1}A^{-1}\| \\ &\leqslant \|(I + A^{-1}E)^{-1}\|_2 \|A^{-1}\|. \end{aligned} \tag{1.16}$$

利用 $\| \ \|_2$ 是相容范数，并且 $\|I\|_2 = 1$，可知（见第二章不等式(2.25)）

$$\begin{aligned} \|(I + A^{-1}E)^{-1}\|_2 &\leqslant (1 - \|A^{-1}E\|_2)^{-1} \leqslant \frac{1}{1 - \|A^{-1}\|_2 \|E\|_2} \\ &= \frac{1}{1 - \|A\| \|A^{-1}\|_2 \frac{\|E\|_2}{\|A\|}} = \frac{1}{\gamma}. \end{aligned} \tag{1.17}$$

将(1.17)代入(1.16)式的右端，便得到不等式(1.12).

2）根据(1.10)和(1.12)，有

$$\begin{aligned} \|B^{-1} - A^{-1}\| &\leqslant \|A^{-1}\|_2 \|A - B\|_2 \|B^{-1}\| \\ &\leqslant \|A\| \|A^{-1}\|_2 \frac{\|E\|_2}{\|A\|} \frac{\|A^{-1}\|}{\gamma}. \end{aligned}$$

由此立即导出不等式(1.13). □

注 1.1. 在定理 1.1 中，对矩阵 E 并没有提出限制条件(1.11)，但结论中需要用到 $\|B^{-1}\|$；在定理 1.2 中，对 E 加了限制条件(1.11)，而在结论中不需要 $\|B^{-1}\|$，这是定理 1.2 的优越之处.

注 1.2. 利用证明上述定理 1.2 的方法，可证(本节习题 1)：如果$\|\ \|$是任一相容的矩阵范数，满足 $\|I\|=1$，则对于任一非奇异阵 A，当 $\|A^{-1}\|\|E\|<1$ 时，$B=A+E$ 必为非奇异阵，并且有估计式

$$\frac{\|B^{-1}-A^{-1}\|}{\|A\|}\leqslant\frac{\kappa(A)\dfrac{\|E\|}{\|A\|}}{1-\kappa(A)\dfrac{\|E\|}{\|A\|}},\tag{1.18}$$

其中

$$\kappa(A)=\|A\|\|A^{-1}\|.\tag{1.19}$$

1.2 线性方程组解的扰动界限

定理 1.3. 设 $A\in\mathrm{C}^{n\times n}$ 是非奇异阵，$\boldsymbol{b}\in\mathrm{C}^n$，$\boldsymbol{x}$ 是方程 $A\boldsymbol{x}=\boldsymbol{b}$ 的解. 又设 $B=A+E$ 满足条件(1.11). 则方程 $B(\boldsymbol{x}+\boldsymbol{h})=\boldsymbol{b}+\boldsymbol{k}$ 有唯一解 $\boldsymbol{x}+\boldsymbol{h}$，并且满足不等式

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|}\leqslant\frac{\kappa}{\gamma}\left(\frac{\|E\|_2}{\|A\|}+\frac{\|\boldsymbol{k}\|}{\|\boldsymbol{b}\|}\right),\tag{1.20}$$

其中 κ 与 γ 分别如(1.14)与(1.15)所示.

证明:

首先由条件 (1.11) 知 B 必为非奇异阵，因此方程组 $B(\boldsymbol{x}+\boldsymbol{h})=\boldsymbol{b}+\boldsymbol{k}$ 有唯一解 $\boldsymbol{x}+\boldsymbol{h}$. 以下证明估计式(1.20).

据 $B(\boldsymbol{x}+\boldsymbol{h})=\boldsymbol{b}+\boldsymbol{k}$，有

$$\begin{aligned}\boldsymbol{h}&=B^{-1}(\boldsymbol{b}+\boldsymbol{k})-\boldsymbol{x}=B^{-1}\boldsymbol{b}+B^{-1}\boldsymbol{k}-\boldsymbol{x}\\&=B^{-1}\boldsymbol{k}-\boldsymbol{x}+A^{-1}[I-(A-B)A^{-1}]^{-1}\boldsymbol{b}.\end{aligned}$$

利用恒等式 $(I-M)^{-1}=I+(I-M)^{-1}M$，得到

$$\begin{aligned}\boldsymbol{h}&=B^{-1}\boldsymbol{k}-\boldsymbol{x}+A^{-1}\{I+[I-(A-B)A^{-1}]^{-1}(A-B)A^{-1}\}\boldsymbol{b}\\&=B^{-1}\boldsymbol{k}-\boldsymbol{x}+A^{-1}\boldsymbol{b}+A^{-1}[I-(A-B)A^{-1}]^{-1}(A-B)A^{-1}\boldsymbol{b}\\&=B^{-1}\boldsymbol{k}+A^{-1}[I-(A-B)A^{-1}]^{-1}(A-B)\boldsymbol{x}\\&=[(B^{-1}-A^{-1})+A^{-1}]\boldsymbol{k}\\&\quad+A^{-1}[I-(A-B)A^{-1}]^{-1}(A-B)\boldsymbol{x}.\end{aligned}\tag{1.21}$$

于是

$$\|\boldsymbol{h}\| \leqslant (\|B^{-1} - A^{-1}\| + \|A^{-1}\|_2)\|\boldsymbol{k}\| + \|A^{-1}\|_2\|[I - (A - B)A^{-1}]^{-1}\|_2\|E\|\|\boldsymbol{x}\|. \quad (1.22)$$

利用 (1.10),

$$\begin{aligned}\|B^{-1} - A^{-1}\| &\leqslant \|A^{-1}\|_2\|A - B\|_2\|B^{-1}\| \\ &= \|A^{-1}\|_2\|E\|_2\|(I + A^{-1}E)^{-1}A^{-1}\| \\ &\leqslant \|A^{-1}\|_2^2\|E\|_2/(1 - \|A^{-1}\|_2\|E\|_2). \quad (1.23)\end{aligned}$$

将(1.23)和(1.17)代入(1.22)式右端，导出

$$\begin{aligned}\|\boldsymbol{h}\| &\leqslant \frac{\|A^{-1}\|_2\|\boldsymbol{k}\|}{1 - \|A^{-1}\|_2\|E\|_2} + \frac{\|A^{-1}\|_2\|E\|_2\|\boldsymbol{x}\|}{\gamma} \\ &= \frac{\|A^{-1}\|_2}{\gamma}(\|\boldsymbol{k}\| + \|E\|_2\|\boldsymbol{x}\|),\end{aligned}$$

从而

$$\begin{aligned}\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|} &\leqslant \frac{\|A^{-1}\|_2}{\gamma}\left(\frac{\|\boldsymbol{k}\|}{\|\boldsymbol{x}\|} + \|E\|_2\right) \\ &= \frac{\|A\|\|A^{-1}\|_2}{\gamma}\left(\frac{\|\boldsymbol{k}\|}{\|A\|\|\boldsymbol{x}\|} + \frac{\|E\|_2}{\|A\|}\right) \\ &\leqslant \frac{\kappa}{\gamma}\left(\frac{\|\boldsymbol{k}\|}{\|\boldsymbol{b}\|} + \frac{\|E\|_2}{\|A\|}\right).\end{aligned}$$

□

注 1.3 估计式 (1.13) 表明，(1.14)式所示的 κ 反映了 A^{-1} 对于 A 的扰动的敏感性；因此 κ 叫做 A 关于矩阵求逆的条件数. 估计式(1.20)表明，κ 还反映了方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的解 $\boldsymbol{x}$ 的相对误差对于 A 和 $\boldsymbol{b}$ 的相对误差的依赖程度；因此，κ 也叫做方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 求解的条件数. 计算可知，例 1.1 中的矩阵 A 的条件数

$$\kappa_2 = \|A\|_2\|A^{-1}\|_2 \approx 8.9443 \times 123.56 \approx 1105.$$

所以，A 和 $\boldsymbol{b}$ 的很小的扰动，可以引起 A 的逆和 $A\boldsymbol{x} = \boldsymbol{b}$ 的解的很大的偏差.

因此，在矩阵求逆和方程组求解时，往往需要考虑如何通过相似变换，降低矩阵的条件数(即所谓“平衡”，或“预处理”). 这是一个仍待进一步研究的问题 (参看第三章注 6.1).

习题

1. 设$\| \quad \|$是任一相容的酉不变范数，满足 $\|I\|=1$. 又设 $A \in \mathbf{C}^{n\times n}$ 为非奇异阵，$E \in \mathbf{C}^{n\times n}$ 满足 $\|A^{-1}\|\|E\| < 1$. 试证：$B = A + E$ 必为非奇异阵，并且有估计式(1.18)，其中 $\kappa(A)$ 为(1.19)式所示。

2. 设 $A \in \mathbf{C}^{n\times n}$ 为任一非奇异阵，κ 与 $\kappa(A)$ 分别如(1.14)式与(1.19)式所示. 证明 $\kappa \geqslant 1$ 和 $\kappa(A) \geqslant 1$，并给出 $\kappa = 1$ 的充分与必要条件。

3. 设 $A \in \mathbf{C}^{n\times n}$ 是非奇异阵，$\boldsymbol{x}$ 和 $\tilde{\boldsymbol{x}}$ 分别是 $A\boldsymbol{x} = \boldsymbol{b}$ 的精确解和近似解. 令 $\boldsymbol{r} = \boldsymbol{b} - A\tilde{\boldsymbol{x}}$. 试证：对任一酉不变范数$\| \quad \|$，有不等式

$$\frac{1}{\kappa} \cdot \frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \leqslant \frac{\|\tilde{\boldsymbol{x}} - \boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leqslant \kappa \cdot \frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|},$$

其中 κ 如(1.14)式所示。

§2 广义逆扰动分析

2.1 关于一对投影

在本节和以下几节，将把正交投影算子（见第一章 §7）简单地叫做投影。

第一章 §7 中的 7.3 已经指出，Hermite 幂等阵

$$P_A = AA^{\dagger} \tag{2.1}$$

是到列空间 $R(A)$ 上的投影，有 $R(P_A) = R(A)$; Hermite 幂等阵

$$P_{A^H} = A^{\dagger}A \tag{2.2}$$

是到列空间 $R(A^H)$ 上的投影，有 $R(P_{A^H}) = R(A^H)$.

到 $R(A)$ 的正交补子空间 $R(A)^{\perp}$ 上的投影，将用

$$P_A^{\perp} \equiv I - P_A \tag{2.3}$$

表示；类似地，将用

$$P_{A^H}^{\perp} \equiv I - P_{A^H} \tag{2.4}$$

表示到 $R(A^H)$ 的正交补子空间 $R(A^H)^{\perp}$ 上的投影。

设 $A \in \mathbf{C}^{m\times n}$，$\mathrm{rank}(A) = r$。$E \in \mathbf{C}^{m\times n}$。令

$$B = A + E. \tag{2.5}$$

显然存在酉阵 $U = (U_1, U_2) \in \mathbf{C}^{m\times m}$ 和 $V = (V_1, V_2) \in \mathbf{C}^{n\times n}$，使得

$$U^H A V = \begin{pmatrix} A_{11} & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.6}$$

其中 $U_1 \in \mathbf{C}^{m\times r}$，$V_1 \in \mathbf{C}^{n\times r}$，$A_{11} \in \mathbf{C}^{r\times r}$。有 $R(A) = R(U_1)$，$R(A^H) = R(V_1)$。相应地，记

$$\left.\begin{aligned} U^H E V &= \underset{\quad r \quad\ n-r}{\begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}} \begin{matrix} r \\ m-r, \end{matrix} \\ U^H B V &= \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \underset{\qquad r \qquad\quad n-r}{\begin{pmatrix} A_{11} + E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}} \begin{matrix} r \\ m-r. \end{matrix} \end{aligned}\right\} \tag{2.7}$$

为了简化讨论，今后将假定 A，E 和 B 具有下述约简形式：

$$A = \underset{r \quad n-r}{\begin{pmatrix} A_{11} & 0 \\ 0 & 0 \end{pmatrix}} \begin{matrix} r \\ m-r, \end{matrix} \quad E = \underset{r \quad n-r}{\begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}} \begin{matrix} r \\ m-r, \end{matrix} \tag{2.8}$$

$$\mathrm{rank}(A_{11}) = r$$

和

$$B = A + E = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11} + E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}. \tag{2.9}$$

利用约简形式 (2.8)—(2.9)，可以立即写出

$$A^{\dagger} = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \tag{2.10}$$

和

$$P_A = \begin{pmatrix} I^{(r)} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbf{C}^{m\times m}, \quad P_{A^H} = \begin{pmatrix} I^{(r)} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbf{C}^{n\times n}. \tag{2.11}$$

并且由此可得

$$\|P_A A P_A{}^H\| = \|A_{11}\| \tag{2.12}$$

和

$$\left.\begin{aligned}&\|P_A E P_A{}^H\| = \|E_{11}\|, \quad \|P_A E P_A^{\perp H}\| = \|E_{12}\|,\\ &\|P_A^{\perp} E P_A{}^H\| = \|E_{21}\|, \quad \|P_A^{\perp} E P_A^{\perp H}\| = \|E_{22}\|.\end{aligned}\right\} \tag{2.13}$$

容易看出,上述恒等式(2.12)和(2.13)对于一般形式的 A, E 和 B(见(2.5)—(2.7)式),也是成立的.

在下面的讨论中,将用到一对投影的若干性质.

首先证明一条矩阵分解定理,它是第一章定理 3.5 的推广.

定理 2.1[23]. 设 $X_1 \in \mathbf{C}^{n\times r}$; $Y_1 \in \mathbf{C}^{n\times s}$, $X_1^H X_1 = I^{(r)}$, $Y_1^H Y_1 = I^{(s)}$, $r \leqslant s$. 则必存在 $n\times n$ 酉阵 Q, $r\times r$ 酉阵 U_1 和 $s\times s$ 酉阵 V_1, 使得

$$QX_1U_1 = \begin{pmatrix} I^{(r)} \\ 0 \\ 0 \end{pmatrix}\begin{matrix} r \\ s-r, \\ n-s \end{matrix} \quad QY_1V_1 = \begin{pmatrix} \Gamma & 0 \\ 0 & I \\ \Sigma & 0 \end{pmatrix}\begin{matrix} r \\ s-r, \\ n-s \end{matrix} \tag{2.14}$$
$$\qquad\quad r \qquad\qquad\qquad\qquad\qquad r \quad s-r$$

其中

$$\left.\begin{aligned}&\Gamma = \begin{pmatrix} \Gamma_1 & 0 \\ 0 & I \end{pmatrix}\begin{matrix} r_1 \\ r-r_1, \end{matrix} \quad \Gamma_1 = \mathrm{diag}(\gamma_1, \cdots, \gamma_{r_1}),\\ &\qquad\ r_1 \quad r-r_1 \qquad\qquad 0 \leqslant \gamma_1 \leqslant \cdots \leqslant \gamma_{r_1} < 1,\end{aligned}\right\} \tag{2.15}$$

$$\left.\begin{aligned}&\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{matrix} r_1 \\ n-s-r_1, \end{matrix} \quad \Sigma_1 = \mathrm{diag}(\sigma_1, \cdots, \sigma_{r_1}),\\ &\qquad\ r_1 \quad r-r_1 \qquad\qquad 1 \geqslant \sigma_1 \geqslant \cdots \geqslant \sigma_{r_1} > 0,\end{aligned}\right\} \tag{2.16}$$

并且 σ_i 与 γ_i 满足

$$\gamma_i^2 + \sigma_i^2 = 1, \quad i = 1, \cdots, r_1. \tag{2.17}$$

证明:

1) 构造酉阵 $X = (X_1, X_2) \in \mathbf{C}^{n\times n}$. 则

$$X^H X_1 = \begin{pmatrix} I^{(r)} \\ 0 \end{pmatrix}, \quad X^H Y_1 = \begin{pmatrix} Y_{11} \\ * \end{pmatrix}\begin{matrix} r \\ n-r. \end{matrix} \tag{2.18}$$

然后对 $Y_{11} \in \mathbf{C}^{r\times s}$ 进行奇异值分解:

$$Y_{11}=U_1\begin{pmatrix}\Gamma_1&0&0\\0&I&0\end{pmatrix}V_1^H,\tag{2.19}$$

其中 $U_1\in\mathbf{C}^{r\times r}$ 与 $V_1\in\mathbf{C}^{s\times s}$ 是酉阵，Γ_1 如(2.15)式所示．于是有

$$\begin{pmatrix}U_1^H&0\\0&I\end{pmatrix}X^HX_1U_1=\begin{pmatrix}I^{(r)}\\0\end{pmatrix},$$

$$\begin{pmatrix}U_1^H&0\\0&I\end{pmatrix}X^HY_1V_1=\begin{array}{c}\begin{pmatrix}\Gamma_1&0&0\\0&I&0\\Y_{21}&0&Y_{22}\end{pmatrix}\begin{matrix}r_1\\r-r_1.\\n-r\end{matrix}\\ \begin{matrix}r_1&r-r_1&s-r\end{matrix}\end{array}\tag{2.20}$$

在上式最右端分块矩阵的第(3,2)位置出现一个 $(n-r)\times(r-r_1)$ 零矩阵，是因为 $Y_1^HY_1=I^{(s)}$．

2）考虑矩阵 $Y'=\begin{array}{c}(Y_{21},\ \ Y_{22})\ n-r\\ \begin{matrix}r_1&s-r\end{matrix}\end{array}$．注意到

$$Y'^HY'=\begin{pmatrix}I^{(r_1)}-\Gamma_1^2&0\\0&I^{(s-r)}\end{pmatrix}>0,\tag{2.21}$$

所以必有 $n-r\geqslant r_1+s-r$，即

$$n-s-r_1\geqslant 0.\tag{2.22}$$

同时由(2.21)可知，Y' 的诸列互相正交，并且 Y_{22} 的诸列皆为单位向量，因此存在酉阵 $U_2\in\mathbf{C}^{(n-r)\times(n-r)}$，使得

$$(Y_{21},Y_{22})=U_2\begin{array}{c}\begin{pmatrix}0&I\\\Sigma_1&0\\0&0\end{pmatrix}\begin{matrix}s-r\\r_1\\n-s-r_1\end{matrix}\\ \begin{matrix}r_1&s-r\end{matrix}\end{array};\tag{2.23}$$

于是得到

$$\begin{pmatrix}I&0\\0&U_2^H\end{pmatrix}\begin{pmatrix}U_1^H&0\\0&I\end{pmatrix}X^HX_1U_1=\begin{array}{c}\begin{pmatrix}I&0\\0&I\\0&0\\0&0\\0&0\end{pmatrix}\begin{matrix}r_1\\r-r_1\\s-r\\r_1\\n-s-r_1\end{matrix}\\ \begin{matrix}r_1&r-r_1\end{matrix}\end{array}\tag{2.24}$$

和

$$\begin{pmatrix} I & 0 \\ 0 & U_2^H \end{pmatrix}\begin{pmatrix} U_1^H & 0 \\ 0 & I \end{pmatrix} X^H Y_1 V_1 = \begin{pmatrix} \Gamma_1 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ \Sigma_1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{matrix} r_1 \\ r-r_1 \\ s-r \\ r_1 \\ n-s-r_1 \end{matrix} . \quad (2.25)$$

$$\qquad\qquad\qquad\qquad r_1 \quad r-r_1 \quad s-r$$

令

$$Q = \begin{pmatrix} I & 0 \\ 0 & U_2^H \end{pmatrix}\begin{pmatrix} U_1^H & 0 \\ 0 & I \end{pmatrix} X^H,$$

它显然是 $n \times n$ 酉阵，利用(2.15)—(2.17)，由 (2.24) 和 (2.25) 可立即写出分解式 (2.14). □

注 2.1. 第一章定理 3.5 的结论是上述定理 2.1 的特例. 事实上，可在定理 2.1 中取 $s = r = l$, 当 $2l \leqslant n$ 时，(2.25) 式的右端即为

$$\begin{pmatrix} \Gamma_1 & 0 \\ 0 & I \\ \Sigma_1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{matrix} r_1 \\ l-r_1 \\ r_1 \\ l-r_1 \\ n-2l \end{matrix} = \begin{pmatrix} \Gamma' \\ \Sigma' \\ 0 \end{pmatrix} \begin{matrix} l \\ l \\ n-2l, \end{matrix}$$

$$r_1 \quad l-r_1$$

上式右端的 Γ' 与 Σ'，就是第一章(3.10)式右端的 Γ 与 Σ；当 $2l \geqslant n$ 时，(2.25) 式的右端即为

$$\begin{pmatrix} \Gamma_1 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ \Sigma_1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{matrix} r_1 \\ n-l-r_1 \\ 2l-n \\ r_1 \\ n-l-r_1 \end{matrix} = \begin{pmatrix} \Gamma' & 0 \\ 0 & I \\ \Sigma' & 0 \end{pmatrix} \begin{matrix} n-l \\ 2l-n \\ n-l \end{matrix},$$

$$r_1 \quad n-l-r_1 \quad 2l-n \qquad\qquad n-l \quad 2l-n$$

上式右端的 Γ' 与 Σ'，就是第一章(3.13)式右端的 Γ 与 Σ.

利用定理 2.1，可以导出关于一对投影的标准形定理.

定理 2.2[23]. 设 $A, B \in \mathbf{C}^{m\times n}$, $\mathrm{rank}(A)=r$, $\mathrm{rank}(B)=s$, $r\leqslant s$. 则存在酉阵 $Q\in \mathbf{C}^{m\times m}$, 使得

$$QP_AQ^H=\begin{pmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{matrix} r \\ s-r \\ m-s \end{matrix} \qquad (2.26)$$
$$\qquad\qquad r \quad s-r \quad m-s$$

和

$$QP_BQ^H=\begin{pmatrix} \Gamma^2 & 0 & \Gamma\Sigma \\ 0 & I & 0 \\ \Sigma\Gamma & 0 & \Sigma^2 \end{pmatrix}\begin{matrix} r \\ s-r, \\ m-s \end{matrix} \qquad (2.27)$$
$$\qquad\qquad r \quad s-r \quad m-s$$

其中

$$\Gamma=\begin{pmatrix} \Gamma_1 & 0 \\ 0 & I \end{pmatrix}\begin{matrix} r_1 \\ r-r_1 \end{matrix},$$
$$\qquad r_1 \quad r-r_1 \qquad (2.28)$$
$$\Sigma=\begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}\begin{matrix} r_1 \\ m-s-r_1 \end{matrix},$$
$$\qquad r_1 \quad r-r_1$$

Γ_1 和 Σ_1 分别如 (2.15) 和 (2.16) 所示.

证明:

据题设, A 和 B 存在满秩分解

$$A=X_1A_1,\ X_1^HX_1=I^{(r)},\ A_1\in \mathbf{C}^{r\times n},\ \mathrm{rank}(A_1)=r \qquad (2.29)$$

和

$$B=Y_1B_1,\ Y_1^HY_1=I^{(s)},\ B_1\in \mathbf{C}^{s\times n},\ \mathrm{rank}(B_1)=s. \qquad (2.30)$$

由 (2.1) 和第一章的**定理** 6.2 可知

$$P_A=(X_1A_1)(X_1A_1)^\dagger=X_1X_1^H,$$
$$P_B=(Y_1B_1)(Y_1B_1)^\dagger=Y_1Y_1^H. \qquad (2.31)$$

利用**定理** 2.1 的 (2.14)—(2.17) 式，立即得出(2.26)和 (2.27). □

注 2.2. 把 (2.28) 代入 (2.26)和(2.27)，得到

$$
\left.\begin{aligned}
QP_AQ^H &= \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}
\begin{matrix} r_1 \\ r-r_1 \\ s-r \\ r_1 \\ m-s-r_1 \end{matrix} \\
&\quad\;\; \begin{matrix} r_1 & r-r_1 & s-r & r_1 & m-s-r_1 \end{matrix} \\
\text{和} \quad QP_BQ^H &= \begin{pmatrix} \Gamma_1^2 & 0 & 0 & \Gamma_1\Sigma_1 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ \Sigma_1\Gamma_1 & 0 & 0 & \Sigma_1^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}
\begin{matrix} r_1 \\ r-r_1 \\ s-r \\ r_1 \\ m-s-r_1 \end{matrix} . \\
&\quad\;\; \begin{matrix} r_1 & r-r_1 & s-r & r_1 & m-s-r_1 \end{matrix}
\end{aligned}\right\} \tag{2.32}
$$

定理 2.3. 设 $A, B \in \mathbf{C}^{m\times n}$. 下述结论成立:

1) 若 $\text{rank}(A) = \text{rank}(B)$, 则 $P_AP_B^\perp$ 与 $P_BP_A^\perp$ 具有相同的奇异值，因而

$$\|P_AP_B^\perp\| = \|P_BP_A^\perp\|. \tag{2.33}$$

此外，有

$$\|P_B - P_A\|_2 = \|P_AP_B^\perp\|_2 = \|P_BP_A^\perp\|_2. \tag{2.34}$$

2) 若 $\|P_B - P_A\|_2 < 1$, 则 $\text{rank}(A) = \text{rank}(B)$.

3) 若 $\text{rank}(B) \geqslant \text{rank}(A)$, 则

$$\|P_BP_A^\perp\| \geqslant \|P_B^\perp P_A\|. \tag{2.35}$$

证明:

1) 根据定理 2.2, 利用标准形 (2.32), 当 $\text{rank}(A)=\text{rank}(B)$ 时，有

$$
\left.\begin{aligned}
P_AP_B^\perp &= Q^H\begin{pmatrix} \Sigma_1^2 & 0 & -\Gamma_1\Sigma_1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}Q \\
P_BP_A^\perp &= Q^H\begin{pmatrix} 0 & 0 & \Gamma_1\Sigma_1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \Sigma_1^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}Q
\end{aligned}\right\} \tag{2.36}
$$

和

$$P_B - P_A = Q^H \begin{pmatrix} -\Sigma_1^2 & 0 & \Gamma_1\Sigma_1 & 0 \\ 0 & 0 & 0 & 0 \\ \Sigma_1\Gamma_1 & 0 & \Sigma_1^2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} Q. \qquad (2.37)$$

由此立即得出等式(2.33)和

$$\|P_B - P_A\|_2 = \|P_A P_B^\perp\|_2 = \|P_B P_A^\perp\|_2 = \sigma_1.$$

2) 设 $\operatorname{rank}(A) = r$, $\operatorname{rank}(B) = s$, 无妨假定 $r \leqslant s$. 根据定理 2.2,有

$$\|P_B - P_A\|_2 = \left\| \begin{pmatrix} -\Sigma_1^2 & 0 & 0 & \Gamma_1\Sigma_1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I^{(s-r)} & 0 & 0 \\ \Sigma_1\Gamma_1 & 0 & 0 & \Sigma_1^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \right\|_2$$

$$= \begin{cases} 1 & \text{当 } s > r \text{ 时} \\ \sigma_1 \leqslant 1 & \text{当 } s = r \text{ 时}. \end{cases} \qquad (2.38)$$

从(2.38)式可以看出,如果 $\operatorname{rank}(A) \neq \operatorname{rank}(B)$ (无妨假定 $\operatorname{rank}(B) > \operatorname{rank}(A)$),则必有 $\|P_B - P_A\|_2 = 1$. 因此结论 2) 成立.

3) 当 $\operatorname{rank}(B) = s \geqslant r = \operatorname{rank}(A)$ 时,根据定理 2.2,有

$$P_B P_A^\perp = Q^H \begin{pmatrix} 0 & 0 & 0 & \Gamma_1\Sigma_1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I^{(s-r)} & 0 & 0 \\ 0 & 0 & 0 & \Sigma_1^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} Q,$$

$$P_B^\perp P_A = Q^H \begin{pmatrix} \Sigma_1^2 & 0 & 0 & -\Gamma_1\Sigma_1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} Q, \qquad (2.39)$$

因此

$$\left.\begin{aligned}&\sigma(P_BP_A^\perp)=\{\underbrace{1,\cdots,1}_{s-r},\sigma_1,\cdots,\sigma_{r_1},\underbrace{0,\cdots,0}_{m-(s-r)-r_1}\}\\&\sigma(P_B^\perp P_A)=\{\sigma_1,\cdots,\sigma_{r_1},\underbrace{0,\cdots,0}_{m-r_1}\}.\end{aligned}\right\}\tag{2.40}$$

利用第二章定理 3.3，立即得到不等式(2.35). □

下述定理表明,对于 $B=A+E$，可以借助于 $\|E\|$ 去估计 $\|P_BP_A^\perp\|$。

定理 2.4. 设 $A\in\mathbb{C}^{m\times n}$，$B=A+E$. 则

$$P_BP_A^\perp=(B^\dagger)^HP_{B^H}E^HP_A^\perp.\tag{2.41}$$

因此，

$$\|P_BP_A^\perp\|\leqslant\|B^\dagger\|_2\|E\|,\tag{2.42}$$

并且当 $\mathrm{rank}(A)=\mathrm{rank}(B)$ 时,有

$$\|P_BP_A^\perp\|\leqslant\min\{\|B^\dagger\|_2,\|A^\dagger\|_2\}\|E\|.\tag{2.43}$$

证明:

利用(2.1)式和 $A^HP_A^\perp=0$，可得

$$\begin{aligned}P_BP_A^\perp&=P_B^HP_A^\perp=(B^\dagger)^HB^HP_A^\perp=(B^\dagger)^H(A+E)^HP_A^\perp\\&=(B^\dagger)^HE^HP_A^\perp=(B^\dagger)^HB^H(B^\dagger)^HE^HP_A^\perp\\&=(B^\dagger)^HP_{B^H}E^HP_A^\perp,\end{aligned}$$

即(2.41)式成立. 由此可得不等式(2.42). 再利用等式(2.33)，便得出估计式(2.43). □

2.2 锐角扰动

对于(2.14)—(2.17)中的 γ_i 与 σ_i，可以记

$$\left.\begin{aligned}&\gamma_i=\cos\theta_i,\sigma_i=\sin\theta_i,i=1,\cdots,r;\\&\frac{\pi}{2}\geqslant\theta_1\geqslant\cdots\geqslant\theta_{r_1}>\theta_{r_1+1}=\cdots=\theta_r=0.\end{aligned}\right.\tag{2.44}$$

这些 $\theta_1\cdots,\theta_r$ 表示经过一个适当的酉变换，并且分别在 $R(X_1)$ 和 $R(Y_1)$ 中适当选取基底向量之后，$R(X_1)$ 的 r 个基底向量和 $R(Y_1)$ 的前 r 个相对应的基底向量之间的夹角.

定理 2.3 的第 2) 条结论已经指出，当 $\|P_B - P_A\|_2 < 1$ 时，必有 $\operatorname{rank}(A) = \operatorname{rank}(B)$，并且根据 (2.38)，

$$\|P_B - P_A\|_2 = \sin\theta_1 < 1,$$

即

$$0 \leqslant \theta_i < \frac{\pi}{2}, \quad i = 1, \cdots, r.$$

此外，根据 (2.29) 和 (2.30)，$R(X_1) = R(A)$，$R(Y_1) = R(B)$；所以，当 $\|P_B - P_A\|_2 < 1$ 时，经过适当的酉变换，并且分别在 $R(A)$ 和 $R(B)$ 中适当选取基底向量之后，$R(A)$ 的 r 个基底向量和 $R(B)$ 的 r 个相对应的基底向量之间的夹角为锐角.

因此，引进下述定义.

定义 3.1. 设 $A, B \in \mathbb{C}^{m\times n}$. 如果 $\|P_B - P_A\| < 1$，则称 $R(A)$ 与 $R(B)$ 互成锐角. 如果 $R(A)$ 与 $R(B)$ 互成锐角，同时 $R(A^H)$ 与 $R(B^H)$ 也互成锐角，则称矩阵 A 与 B 互成锐角，或者说，B 是 A 的锐角扰动.

定理 2.5. $R(A)$ 与 $R(B)$ 互成锐角的必要与充分条件是在 $R(A)$ 内不存在向量与 $R(B)$ 正交，同时在 $R(B)$ 内不存在向量与 $R(A)$ 正交.

证明:

必要性. 如果存在向量 $\boldsymbol{x} \neq 0$，满足 $P_A\boldsymbol{x} = \boldsymbol{x}$ 和 $P_B\boldsymbol{x} = 0$，则有 $(P_B - P_A)\boldsymbol{x} = -\boldsymbol{x}$，从而 $\|P_B - P_A\|_2 \geqslant 1$；这与 $\|P_B - P_A\|_2 < 1$ 矛盾.

充分性. 如果 $\|P_B - P_A\|_2 = 1$，则因 $P_B - P_A$ 为 Hermite 阵，必存在向量 $\boldsymbol{x} \neq 0$，使得 $(P_B - P_A)\boldsymbol{x} = \boldsymbol{x}$ 或 $(P_A - P_B)\boldsymbol{x} = \boldsymbol{x}$. 无妨假定 $(P_B - P_A)\boldsymbol{x} = \boldsymbol{x}$.

若 $P_A\boldsymbol{x} = 0$，则有 $P_B\boldsymbol{x} = \boldsymbol{x}$，这表示存在与 $R(A)$ 正交的向量 $\boldsymbol{x} \in R(B)$；显然与题设矛盾. 假若 $P_A\boldsymbol{x} \neq 0$，则由 $P_A\boldsymbol{x} = -(I - P_B)\boldsymbol{x}$ 可知，$P_B(P_A\boldsymbol{x}) = 0$，这表示存在非零向量 $P_A\boldsymbol{x} \in R(A)$，使得 $P_A\boldsymbol{x} \in R(B)^\perp$；显然也与题设矛盾. □

下述定理给出了 A 与 B 互成锐角的必要与充分条件.

定理 2.6.[157] 设 $A, B \in \mathbb{C}^{m\times n}$. 则 A 与 B 互成锐角的必要与充分条件是

$$\operatorname{rank}(A) = \operatorname{rank}(B) = \operatorname{rank}(P_A B P_{A^H}). \tag{2.45}$$

证明:

不失一般性，在证明时可以使用 A 与 B 的约简形式 (2.8) 与 (2.9). 充分性. 由(2.45)和(2.11)可知, $\operatorname{rank}(A_{11}) = \operatorname{rank}(B) = \operatorname{rank}(B_{11}) = r$. 于是有

$$\left.\begin{aligned} R(B) &= R\left(\begin{pmatrix} B_{11} \\ E_{21} \end{pmatrix}\right) = R\left(\begin{pmatrix} I^{(r)} \\ E_{21}B_{11}^{-1} \end{pmatrix}\right), \\ R(A) &= R\left(\begin{pmatrix} A_{11} \\ 0 \end{pmatrix}\right) = R\left(\begin{pmatrix} I^{(r)} \\ 0 \end{pmatrix}\right) \end{aligned}\right\} \tag{2.46}$$

和

$$P_A = P\begin{pmatrix} I^{(r)} \\ 0 \end{pmatrix}, \quad P_B = P\begin{pmatrix} I \\ E_{21}B_{11}^{-1} \end{pmatrix}. \tag{2.47}$$

应用第 2 章定理 4.2，得到

$$\begin{aligned} \|P_B - P_A\|_2^2 &= \left\| \sin\Theta\left(\begin{pmatrix} I \\ E_{21}B_{11}^{-1} \end{pmatrix}, \begin{pmatrix} I^{(r)} \\ 0 \end{pmatrix}\right)\right\|_2^2 \\ &= \|I - (I + B_{11}^{-H}E_{21}^H E_{21} B_{11}^{-1})^{-1}\|_2 \\ &= \|B_{11}^{-H}E_{21}^H E_{21}B_{11}^{-1}(I + B_{11}^{-H}E_{21}^H E_{21}B_{11}^{-1})^{-1}\|_2. \end{aligned} \tag{2.48}$$

设 $E_{21}B_{11}^{-1}$ 的奇异值分解为 $E_{21}B_{11}^{-1} = U\Sigma V^H$, 其中 U 与 V 为酉阵, $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \cdots), \sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant 0$; 代入 (2.48), 可知

$$\|P_B - P_A\|_2 = \|\Sigma^T\Sigma(I + \Sigma^T\Sigma)^{-1}\|_2^{\frac{1}{2}} = \frac{\sigma_1}{\sqrt{1+\sigma_1^2}} < 1.$$

同理可得 $\|P_{B^H} - P_{A^H}\|_2 < 1$.

必要性. 设 A 与 B 互成锐角. 根据定理 2.3 之 2), 应有 $\operatorname{rank}(A) = \operatorname{rank}(B) \geqslant \operatorname{rank}(B_{11})$. 假若 $\operatorname{rank}(B_{11}) < \operatorname{rank}(A)$, 即 B_{11} 为奇异阵,则存在单位向量 $\boldsymbol{p}$ 与 $\boldsymbol{q}$,使得 $\boldsymbol{p}^H B_{11} = 0, B_{11}\boldsymbol{q} = 0$. 构造酉阵 $P = (\boldsymbol{p}, P_1)$ 与 $Q = (\boldsymbol{q}, Q_1)$. 无妨考虑

$$A = \begin{pmatrix} P^H A_{11} Q & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} P^H B_{11} Q & P^H E_{12} \\ E_{21} Q & E_{22} \end{pmatrix},$$

其中 $P^H B_{11} Q$ 的第 1 行与第 1 列的元素均为零. 如果 $E_{21}\boldsymbol{q} \neq 0$，则存在与 $R(A)$ 正交的非零向量 $\begin{pmatrix} 0 \\ E_{21}\boldsymbol{q} \end{pmatrix} \in R(B)$; 如果 $\boldsymbol{p}^H E_{12} \neq 0$，则存在与 $R(A^H)$ 正交的非零向量 $\begin{pmatrix} 0 \\ E_{12}^H \boldsymbol{p} \end{pmatrix} \in R(A^H)$. 如果 $E_{21}\boldsymbol{q} = 0$ 和 $\boldsymbol{p}^H E_{12} = 0$，则向量 $\boldsymbol{e}_1 \in R(A)$，并且 $\boldsymbol{e}_1 \in R(A^H)$，同时 $\boldsymbol{e}_1$ 与 $R(B)$ 正交，并且与 $R(B^H)$ 正交，总之，所有这些，都与题设矛盾. 所以 B_{11} 必为非奇异阵，即 $\operatorname{rank}(A) = \operatorname{rank}(B) = \operatorname{rank}(B_{11})$. □

2.3 广义逆的扰动界限

例 2.1. $A = \begin{pmatrix} \alpha & 0 \\ 0 & 0 \end{pmatrix}$, $\alpha > 0$; $E = \begin{pmatrix} 0 & \varepsilon \\ \varepsilon & 0 \end{pmatrix}$,

$$B = A + E = \begin{pmatrix} \alpha & \varepsilon \\ \varepsilon & 0 \end{pmatrix}, \varepsilon > 0.$$

显然有

$$A^\dagger = \begin{pmatrix} \frac{1}{\alpha} & 0 \\ 0 & 0 \end{pmatrix}, \quad B^\dagger = B^{-1} = \begin{pmatrix} 0 & \frac{1}{\varepsilon} \\ \frac{1}{\varepsilon} & -\frac{\alpha}{\varepsilon^2} \end{pmatrix}$$

和

$$B^\dagger - A^\dagger = \begin{pmatrix} -\frac{1}{\alpha} & \frac{1}{\varepsilon} \\ \frac{1}{\varepsilon} & -\frac{\alpha}{\varepsilon^2} \end{pmatrix} = \frac{1}{\varepsilon}\begin{pmatrix} -\frac{\varepsilon}{\alpha} & 1 \\ 1 & -\frac{\alpha}{\varepsilon} \end{pmatrix};$$

从而得到

$$\|B^\dagger - A^\dagger\|_2 = \frac{1}{\varepsilon}\left\|\begin{pmatrix} -\frac{\varepsilon}{\alpha} & 1 \\ 1 & -\frac{\alpha}{\varepsilon} \end{pmatrix}\right\|_2$$

$$= \frac{1}{\varepsilon}\,\frac{\varepsilon^2 + \alpha^2}{\alpha\varepsilon} \geqslant \frac{2}{\varepsilon} = \frac{2}{\|E\|_2}$$

和

$$\|B^{\dagger}\|_2 = \frac{1}{\varepsilon}\left\|\begin{pmatrix} 0 & 1 \\ 1 & -\frac{\alpha}{\varepsilon} \end{pmatrix}\right\|_2 = \frac{1}{\varepsilon}\cdot\frac{\alpha+\sqrt{\alpha^2+4\varepsilon^2}}{2\varepsilon}$$

$$\geqslant \frac{1}{\varepsilon} = \frac{1}{\|E\|_2}.$$

上例表明，一般说来，广义逆是不连续的.

在讨论 $\|B^{\dagger}-A^{\dagger}\|$ 的上界之前，先论述一条由例 2.1 所引伸出的一般性命题.

定理 2.7. 设 $A\in\mathbf{C}^{m\times n}$，$B=A+E$. 如果 A 与 B 不成锐角，则

$$\|B^{\dagger}-A^{\dagger}\|_2 \geqslant \frac{1}{\|E\|_2}; \tag{2.49}$$

并且当 $\mathrm{rank}(B)\geqslant\mathrm{rank}(A)$ 时，有

$$\|B^{\dagger}\|_2 \geqslant 1/\|E\|_2. \tag{2.50}$$

证明:

不妨设 $\mathrm{rank}(B)\geqslant\mathrm{rank}(A)$. 因 A 与 B 不成锐角，所以有 $\|P_B-P_A\|_2=1$ 或 $\|P_{B^H}-P_{A^H}\|_2=1$. 以下设 $\|P_B-P_A\|_2=1$（当 $\|P_{B^H}-P_{A^H}\|_2=1$ 时，可同理讨论之）.

利用 A 与 B 的满秩分解(2.29) 与 (2.30)，并考虑到 (2.24)，(2.25)和(2.38)，容易看出，必存在向量 $\boldsymbol{y}\in R(B)$，$\|\boldsymbol{y}\|_2=1$，使得 $\boldsymbol{y}\in R(A)^{\perp}$. 于是

$$\begin{aligned}1 &= \boldsymbol{y}^H\boldsymbol{y} = \boldsymbol{y}^H P_B\boldsymbol{y} = \boldsymbol{y}^H BB^{\dagger}\boldsymbol{y} = \boldsymbol{y}^H(A+E)B^{\dagger}\boldsymbol{y} \\ &= \boldsymbol{y}^H EB^{\dagger}\boldsymbol{y} \leqslant \|E\|_2\|B^{\dagger}\boldsymbol{y}\|_2 \leqslant \|E\|_2\|B^{\dagger}\|_2,\end{aligned}$$

因此 (2.50) 成立. 再由 $A^{\dagger}\boldsymbol{y}=A^{\dagger}P_A\boldsymbol{y}=0$ 可知

$$\frac{1}{\|E\|_2} \leqslant \|B^{\dagger}\boldsymbol{y}\|_2 \leqslant \|(B^{\dagger}-A^{\dagger})\boldsymbol{y}\|_2 \leqslant \|B^{\dagger}-A^{\dagger}\|_2. \qquad \square$$

定理 2.7 表明，广义逆一般不是矩阵元素的连续函数；此外，它还表明，如果 A 与 B 不成锐角，并且 $\mathrm{rank}(A)=\mathrm{rank}(B)$，则

$\|A^\dagger\|_2$与$\|B^\dagger\|_2$都不小于 $1/\|E\|_2$.

尽管如此,仍然可以求出 $\|B^\dagger - A^\dagger\|$ 的上界(见下面的定理2.9);当然,随着 B 趋向 A,这些上界一般不再保持有界.

在估计 $\|B^\dagger - A^\dagger\|$ 的上界时,需要用到下述分解定理.

定理 2.8. 设 $A \in \mathbf{C}^{m\times n}, B = A + E$. 则 $B^\dagger - A^\dagger$ 有分解式

$$B^\dagger - A^\dagger = -B^\dagger E A^\dagger + B^\dagger P_A^\perp - P_{B^H}^\perp A^\dagger, \tag{2.51}$$

$$B^\dagger - A^\dagger = -B^\dagger P_B E P_{A^H} A^\dagger + B^\dagger P_B P_A^\perp - P_{B^H}^\perp P_{A^H} A^\dagger \tag{2.52}$$

和

$$B^\dagger - A^\dagger = -B^\dagger P_B E P_{A^H} A^\dagger + (B^H B)^\dagger P_{B^H} E^H P_A^\perp - P_{B^H}^\perp E P_A (AA^H)^\dagger. \tag{2.53}$$

证明:

将 $E = B - A$, $P_A = AA^\dagger$, $P_{A^H} = A^\dagger A$, $P_B = BB^\dagger$ 与 $P_{B^H} = B^\dagger B$ 代入(2.52)式右端,得到

$$\begin{aligned}
&-B^\dagger BB^\dagger(B - A)A^\dagger AA^\dagger + B^\dagger BB^\dagger(I - AA^\dagger) \\
&\qquad - (I - B^\dagger B)A^\dagger AA^\dagger \\
&= -B^\dagger BA^\dagger + B^\dagger AA^\dagger + B^\dagger - B^\dagger AA^\dagger - A^\dagger + B^\dagger BA^\dagger \\
&= B^\dagger - A^\dagger,
\end{aligned}$$

即恒等式(2.52)成立. 同理可证(2.51)和(2.53)(本节习题1). □

注 2.3. 在(2.51)和(2.52)中,将 A 与 B 对换,可得恒等式

$$B^\dagger - A^\dagger = -A^\dagger E B^\dagger - A^\dagger P_B^\perp + P_{A^H}^\perp B^\dagger \tag{2.54}$$

和

$$B^\dagger - A^\dagger = -A^\dagger P_A E P_{B^H} B^\dagger - A^\dagger P_A P_{B^H}^\perp + P_{A^H}^\perp P_{B^H} B^\dagger. \tag{2.55}$$

同样地,可在(2.53)中将 A 与 B 对换,从而将恒等式变形.

以下分三种情形,讨论 $\|B^\dagger - A^\dagger\|$ 的上界.

2.3.1 一般情形

定理 2.9.[183] 设 $A \in \mathbf{C}^{m\times n}$, $B = A + E$. 则

$$\|B^\dagger - A^\dagger\| \leqslant \mu \max\{\|A^\dagger\|_2^2, \|B^\dagger\|_2^2\}\|E\|, \tag{2.56}$$

其中 μ 由下表给出:

$\| \ \|$	任一范数	谱范数	Frobenius 范数
μ	3	$\frac{1+\sqrt{5}}{2}$	$\sqrt{2}$

证明:

1) 设$\| \ \|$是任一酉不变范数. 由(2.53)可得

$$
\begin{aligned}
\|B^\dagger - A^\dagger\| &\leqslant \|B^\dagger\|_2 \|P_B\|_2 \|E\| \|P_{A^H}\|_2 \|A^\dagger\|_2 \\
&\quad + \|(B^H B)^\dagger\|_2 \|P_{B^H}\|_2 \|E\| \|P_A^\perp\|_2 \\
&\quad + \|P_{B^H}^\perp\|_2 \|E\| \|P_A\|_2 \|(AA^H)^\dagger\|_2 \\
&\leqslant (\|A^\dagger\|_2 \|B^\dagger\|_2 + \|B^\dagger\|_2^2 + \|A^\dagger\|_2^2) \|E\| \\
&\leqslant 3\max\{\|A^\dagger\|_2^2, \|B^\dagger\|_2^2\} \|E\|.
\end{aligned}
$$

2) 取$\| \ \| = \| \ \|_2$. 当 $\operatorname{rank}(B) \leqslant \operatorname{rank}(A)$ 时,由(2.51)式可知,对于任一向量 $\boldsymbol{u} \in \mathbf{C}^m$, $\|\boldsymbol{u}\|_2 = 1$, 有

$$
\begin{aligned}
\|(B^\dagger - A^\dagger)\boldsymbol{u}\|_2^2 &= \|-B^\dagger E A^\dagger \boldsymbol{u} + B^\dagger P_A^\perp \boldsymbol{u}\|_2^2 + \|P_{B^H}^\perp A^\dagger \boldsymbol{u}\|_2^2 \\
&= \|-B^\dagger E A^\dagger P_A \boldsymbol{u} + B^\dagger P_A^\perp P_A^\perp \boldsymbol{u}\|_2^2 + \|P_{B^H}^\perp A^\dagger P_A \boldsymbol{u}\|_2^2 \\
&\leqslant (\|B^\dagger E A^\dagger\|_2 \|P_A \boldsymbol{u}\|_2 + \|B^\dagger P_A^\perp\|_2 \|P_A^\perp \boldsymbol{u}\|_2)^2 \\
&\quad + \|P_{B^H}^\perp A^\dagger\|_2^2 \|P_A \boldsymbol{u}\|_2^2. \qquad (2.57)
\end{aligned}
$$

令

$$
\alpha_1 = \|B^\dagger E A^\dagger\|_2, \ \alpha_2 = \|B^\dagger P_A^\perp\|_2, \ \alpha_3 = \|P_{B^H}^\perp A^\dagger\|_2 \qquad (2.58)
$$

和

$$
\cos\theta = \|P_A \boldsymbol{u}\|_2 \geqslant 0, \quad \sin\theta = \|P_A^\perp \boldsymbol{u}\|_2 \geqslant 0. \qquad (2.59)
$$

代入(2.57), 得到

$$
\begin{aligned}
\|(B^\dagger - A^\dagger)\boldsymbol{u}\|_2^2 &\leqslant (\alpha_1 \cos\theta + \alpha_2 \sin\theta)^2 + \alpha_3^2 \cos^2\theta \\
&\leqslant \max_{0 \leqslant \theta \leqslant \frac{\pi}{2}} [(\alpha_1 \cos\theta + \alpha_2 \sin\theta)^2 + \alpha_3^2 \cos^2\theta] \\
&= \frac{1}{2} \{\alpha_1^2 + \alpha_2^2 + \alpha_3^2 \\
&\quad + [(\alpha_1^2 + \alpha_2^2 + \alpha_3^2)^2 - 4\alpha_2^2 \alpha_3^2]^{\frac{1}{2}}\} \qquad \text{(本节习题 2)} \\
&\leqslant \frac{3+\sqrt{5}}{2} \max\{\alpha_1^2, \alpha_2^2, \alpha_3^2\}
\end{aligned}
$$

$$= \left(\frac{1+\sqrt{5}}{2}\right)^2 \max\{\alpha_1^2, \alpha_2^2, \alpha_3^2\}. \tag{2.60}$$

显然有

$$\alpha_1 \leqslant \|A^+\|_2 \|B^+\|_2 \|E\|_2; \tag{2.61}$$

利用定理 2.3 之 3) 和 $P_B^\perp B = 0$，有

$$\alpha_2 = \|B^\dagger P_B P_A^\perp\|_2 \leqslant \|B^\dagger\|_2 \|P_B P_A^\perp\|_2 \leqslant \|B^\dagger\|_2 \|P_B^\perp P_A\|_2$$
$$= \|B^\dagger\|_2 \|P_B^\perp (B - A) A^\dagger\|_2 \leqslant \|A^\dagger\|_2 \|B^\dagger\|_2 \|E\|_2; \tag{2.62}$$

此外,利用 $P_{B^H} = B^H B^{H\dagger}$ 和 $A^\dagger A = (A^\dagger A)^H = A^H A^{\dagger H}$ 可得

$$\alpha_3 = \|(I - B^H B^{H\dagger}) A^H A^{\dagger H} A^\dagger\|_2$$
$$= \|(I - B^H B^{H\dagger})(A^H - B^H) A^{\dagger H} A^\dagger\|_2$$
$$\leqslant \|A^\dagger\|_2^2 \|E\|_2. \tag{2.63}$$

将 (2.61)—(2.63) 代入 (2.60)，便得到

$$\|B^\dagger - A^\dagger\|_2 \leqslant \frac{1+\sqrt{5}}{2} \max\{\|A^\dagger\|_2^2, \|B^\dagger\|_2^2\} \|E\|_2. \tag{2.64}$$

当 rank$(B) \geqslant$ rank(A) 时,可利用恒等式 (2.54)，同理证明不等式 (2.64)。

3) 取 $\|\ \| = \|\ \|_F$. 当 rank$(B) \leqslant$ rank (A) 时，由 (2.52) 式可知，如果记 $F_1 = -B^\dagger P_B E P_{A^H} A^\dagger$， $F_2 = B^\dagger P_B P_A^\perp$ 和 $F_3 = -P_B^\perp P_{A^H} A^\dagger$，则有

$$\|B^\dagger - A^\dagger\|_F^2 = \|F_1 + F_2\|_F^2 + \|F_3\|_F^2. \tag{2.65}$$

注意,到 $F_1 + F_2 = B^\dagger(-P_B E A^\dagger P_A + P_B P_A^\perp)$，从而

$$\|F_1 + F_2\|_F^2 \leqslant \|B^\dagger\|_2^2 (\|P_B E A^\dagger P_A\|_F^2 + \|P_B P_A^\perp\|_F^2).$$

利用定理 2.3 之 3)，可得

$$\|P_B E A^\dagger P_A\|_F^2 + \|P_B P_A^\perp\|_F^2$$
$$\leqslant \|P_B E A^\dagger\|_F^2 + \|P_B^\perp P_A\|_F^2$$
$$= \|P_B E A^\dagger\|_F^2 + \|P_B^\perp E A^\dagger\|_F^2$$
$$= \|E A^\dagger\|_F^2 \leqslant \|A^\dagger\|_2^2 \|E\|_F^2.$$

因此

$$\|F_1 + F_2\|_F \leqslant \|A^\dagger\|_2 \|B^\dagger\|_2 \|E\|_F. \tag{2.66}$$

此外 (与 (2.63) 式之推导相同),有

$$\|F_3\|_F \leqslant \|A^\dagger\|_2 \|P_B^{\perp H} P_A^H\|_F$$
$$= \|A^\dagger\|_2 \|P_B^{\perp H} E^H A^\dagger\|_F \leqslant \|A^\dagger\|_2^2 \|E\|_F. \qquad (2.67)$$

将 (2.66) 与 (2.67) 代入 (2.65)，便得到

$$\|B^\dagger - A^\dagger\|_F \leqslant \sqrt{2}\, \|A^\dagger\|_2 \max\{\|A^\dagger\|_2, \|B^\dagger\|_2\} \|E\|_F$$
$$\leqslant \sqrt{2} \max\{\|A^\dagger\|_2^2, \|B^\dagger\|_2^2\} \|E\|_F. \qquad (2.68)$$

当 $\mathrm{rank}(B) \geqslant \mathrm{rank}(A)$ 时，可利用恒等式 (2.55)，同理可证不等式 (2.68). □

值得指出的是，即使 $\|E\|$ 很小，也不保证 $\|B^\dagger - A^\dagger\|$ 的上界（见 (2.56) 式右端）很小，因为当 $\|E\|$ 趋向零时，$\|B^\dagger\|_2$ 可能无限制地增长.

2.3.2 $\mathrm{rank}(A) = \mathrm{rank}(B)$ 的情形

当 $\mathrm{rank}(A) = \mathrm{rank}(B)$ 时，定理 2.9 的结论可以在两个方面得到加强. 一方面是用 $\|A^\dagger\|_2 \|B^\dagger\|_2$ 代替 $\max\{\|A^\dagger\|_2^2, \|B^\dagger\|_2^2\}$，另一方面，是对 $\|B^\dagger - A^\dagger\|$ 上界中的常数 μ，区分更多的不同情况，进行更细致的讨论.

定理 2.10.[183] 设 $A \in \mathbf{C}^{m\times n}$, $B = A + E$. 如果 $\mathrm{rank}(A) = \mathrm{rank}(B)$，则

$$\|B^\dagger - A^\dagger\| \leqslant \mu \|A^\dagger\|_2 \|B^\dagger\|_2 \|E\|, \qquad (2.69)$$

其中 μ 由下表给出：

rank \ $\|\ \|$	任一范数	谱范数	Frobenius 范数
$\mathrm{rank}(A) < \min(m,n)$	3	$\frac{1+\sqrt{5}}{2}$	$\sqrt{2}$
$\mathrm{rank}(A) = \min(m,n)$, $m \neq n$	2	$\sqrt{2}$	1
$\mathrm{rank}(A) = m = n$	1	1	1

证明：

设 $\|\ \|$ 是任一酉不变范数. 由 (2.51) 式可得

$$\|B^\dagger - A^\dagger\| \leqslant \|B^\dagger\|_2 \|E\| \|A^\dagger\|_2$$

$$+\|B^\dagger\|_2\|P_BP_A^\perp\| + \|A^\dagger\|_2\|P_B^{\perp H}P_{A^H}\|. \quad (2.70)$$

利用(2.33)，有

$$\|P_B^{\perp H}P_{A^H}\| = \|P_{A^H}P_B^{\perp H}\| = \|P_{B^H}P_A^{\perp H}\|,$$

代入(2.70)，

$$\|B^\dagger - A^\dagger\| \leqslant \|A^\dagger\|_2\|B^\dagger\|_2\|E\| + \|B^\dagger\|_2\|P_BP_A^\perp\| + \|A^\dagger\|_2\|P_{B^H}P_A^{\perp H}\|. \quad (2.71)$$

注意到

$$\|P_BP_A^\perp\| = \begin{cases} 0 & \text{当 } \operatorname{rank}(A) = m \text{ 时} \\ \|B^{+H}(B-A)^H(I - A^{\dagger H}A^H)\| \leqslant \|B^\dagger\|_2\|E\| & \text{当 } \operatorname{rank}(A) < m \text{ 时} \end{cases} \quad (2.72)$$

和

$$\|P_{B^H}P_A^{\perp H}\| = \begin{cases} 0 & \text{当 } \operatorname{rank}(A) = n \text{ 时} \\ \|B^\dagger(B-A)(I - A^\dagger A)\| \leqslant \|B^\dagger\|_2\|E\| & \text{当 } \operatorname{rank}(A) < n \text{ 时}, \end{cases} \quad (2.73)$$

将(2.72)和(2.73)代入(2.71)式右端，便得到估计式(2.69)。

同理，利用定理 2.3 之 1) 和定理 2.9 证明的 2) 与 3)，以及(2.72)—(2.73)，可证对于范数 $\|\ \|_2$ 与 $\|\ \|_F$ 的估计式(2.69)。□

由定理 2.10 可得

推论 2.1 在定理 2.10 的假设条件下，有

$$\frac{\|B^\dagger - A^\dagger\|}{\|B^\dagger\|_2} \leqslant \mu\kappa_\dagger \frac{\|E\|}{\|A\|}, \quad (2.74)$$

其中 $\kappa_\dagger = \|A^\dagger\|_2\|A\|$。

进而，可以得到与矩阵逆的扰动定理（见定理 1.2）相类似的结果：

定理 2.11.[157] 设 $A \in \mathbf{C}^{m\times n}$，$B = A + E$。如果 $\operatorname{rank}(A) = \operatorname{rank}(B)$，并且 $\|A^\dagger\|_2\|E\|_2 < 1$，则有

$$\|B^\dagger\|_2 \leqslant \|A^\dagger\|_2/\gamma_\dagger \quad (2.75)$$

和

$$\frac{\|B^\dagger - A^\dagger\|}{\|A^\dagger\|_2} \leqslant \frac{\mu\kappa_\dagger}{\gamma_\dagger} \frac{\|E\|}{\|A\|}, \quad (2.76)$$

其中

$$\kappa_{\dagger}=\|A^{\dagger}\|_2\|A\|,\ \gamma_{\dagger}=1-\|A^{\dagger}\|_2\|E\|_2. \tag{2.77}$$

证明:

设 $\operatorname{rank}(A)=\operatorname{rank}(B)=r$，$A$的非零奇异值和$B$的非零奇异值分别为 $\{\sigma_i(A)\}_{i=1}^r$ 和 $\{\sigma_i(B)\}_{i=1}^r$，适合 $\sigma_1(A)\geqslant\cdots\geqslant\sigma_r(A)>0$ 和 $\sigma_1(B)\geqslant\cdots\geqslant\sigma_r(B)>0$. 显然有

$$\|A^{\dagger}\|_2=\frac{1}{\sigma_r(A)},\ \|B^{\dagger}\|_2=\frac{1}{\sigma_r(B)}.$$

根据第三章定理 3.11,

$$|\sigma_r(B)-\sigma_r(A)|\leqslant\|E\|_2,$$

因此有

$$\|E\|_2\geqslant\sigma_r(A)-\sigma_r(B)=\frac{1}{\|A^{\dagger}\|_2}-\frac{1}{\|B^{\dagger}\|_2},$$

即不等式 (2.75) 成立.

将不等式(2.75)代入(2.69)式的右端,便得到

$$\frac{\|B^{\dagger}-A^{\dagger}\|}{\|A^{\dagger}\|_2}\leqslant\mu\frac{\|A^{\dagger}\|_2}{\gamma_{\dagger}}\|E\|=\frac{\mu\|A^{\dagger}\|_2\|A\|}{\gamma_{\dagger}}\cdot\frac{\|E\|}{\|A\|}$$

$$=\frac{\mu\kappa_{\dagger}}{\gamma_{\dagger}}\frac{\|E\|}{\|A\|}.\qquad\square$$

不等式 (2.76) 表明，当 $B\to A$ (即 $E\to 0$) 并且保持 $\operatorname{rank}(B)=\operatorname{rank}(A)$ 时,有 $B^{\dagger}\to A^{\dagger}$; 另一方面,如果 $\operatorname{rank}(B)\neq\operatorname{rank}(A)$，则$A$与$B$必不互成锐角,根据定理 2.7 (见(2.49)式),有 $\|B^{\dagger}-A^{\dagger}\|_2\geqslant 1/\|E\|_2$. 因此，这就证明了广义逆的连续性定理，即

定理 2.12.[151] 设 $A,B\in\mathbf{C}^{m\times n}$，其中$B$视为变量. 则

$$\lim_{B\to A}B^{\dagger}=A^{\dagger}$$

的必要与充分条件是当B充分靠近A并且趋向A时,保持

$$\operatorname{rank}(B)=\operatorname{rank}(A).$$

2.3.3 锐角扰动的情形

设 $\mathbf{C}^{k\times r}$ 上的酉不变范数是由 $\mathbb{R}^r$ 上的一个 SG 函数Φ生成

的. 现定义 $\mathbb{C}^{k\times r}$ 上的函数 ψ_Φ:

$$\psi_\Phi(F)=\Phi\left(\frac{\sigma_1(F)}{[1+\sigma_1^2(F)]^{\frac{1}{2}}},\cdots,\frac{\sigma_r(F)}{[1+\sigma_r^2(F)]^{\frac{1}{2}}}\right),\quad F\in\mathbb{C}^{k\times r},\tag{2.78}$$

其中 $\sigma_1(F),\cdots,\sigma_r(F)$ 是 F 的奇异值.

容易看出 ψ_Φ 具有下列性质:

(i) $\psi_\Phi(GF)\leqslant\psi_\Phi(\|G\|_2F)\leqslant\psi_\Phi(\|G\|F)$.

这是因为奇异值

$$\sigma_i(GF)\leqslant\|G\|_2\sigma_i(F)\leqslant\|G\|\sigma_i(F),i=1,\cdots,r.$$

(ii) $\psi_\Phi(\alpha F)\leqslant\alpha\psi_\Phi(F)$, $\forall\alpha\geqslant1$.

这是因为对于 $\sigma\geqslant0$, 有不等式

$$\frac{\alpha\sigma}{(1+\alpha^2\sigma^2)^{1/2}}\leqslant\frac{\alpha\sigma}{(1+\sigma^2)^{1/2}},\ \forall\alpha\geqslant1.$$

(iii) 当 $\|F\|$ 足够小时,有

$$\psi_\Phi(F)=\|F\|+o(\|F\|).$$

(iv) $\psi_\Phi(F)\leqslant\|I^{(r)}\|$.

(v) 对于谱范数,有

$$\psi_2(F)=\frac{\|F\|_2}{(1+\|F\|_2^2)^{\frac{1}{2}}}.$$

引理 2.1. 设 $F\in\mathbb{C}^{k\times r}$. 则有

$$\left\|\begin{pmatrix}I\\F\end{pmatrix}^\dagger\right\|_2\leqslant1\tag{2.79}$$

和

$$\left\|\begin{pmatrix}I\\F\end{pmatrix}^\dagger-(I^{(r)},0)\right\|=\psi_\Phi(F).\tag{2.80}$$

证明:

因为

$$\begin{pmatrix}I\\F\end{pmatrix}^\dagger=(I+F^HF)^{-1}(I,F),\tag{2.81}$$

它的奇异值是

$$\frac{1}{[1+\sigma_i^2(F)]^{\frac{1}{2}}} \leqslant 1, i=1,\cdots,r.$$

所以不等式 (2.79) 成立.

令

$$G=\begin{pmatrix} I \\ F \end{pmatrix}^{\dagger}-(I^{(r)}, 0).$$

计算可知

$$GG^H = I-(I+F^HF)^{-1},$$

所以 G 的奇异值是

$$\frac{\sigma_i(F)}{[1+\sigma_i^2(F)]^{1/2}}, \quad i=1,\cdots,r.$$

由此立即得到 (2.80). □

在下面的讨论中,将使用 A 与 B 的约简形式(2.8)与(2.9).

引理 2.2. 设 $A\in \mathbf{C}_r^{m\times n}$, B 是 A 的锐角扰动. 则

$$B^{\dagger}=(I, F_{12})^{\dagger}B_{11}^{-1}\begin{pmatrix} I \\ F_{21} \end{pmatrix}^{\dagger}, \tag{2.82}$$

其中

$$F_{21}=E_{21}B_{11}^{-1}, \ \ F_{12}=B_{11}^{-1}E_{12}. \tag{2.83}$$

证明:

根据 B 是 A 的锐角扰动, 可知

$$R(B)=R\left(\begin{pmatrix} B_{11} \\ E_{21} \end{pmatrix}\right).$$

因此存在 F_{12}, 使得

$$\begin{aligned} B &= \begin{pmatrix} B_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix} = \begin{pmatrix} B_{11} \\ E_{21} \end{pmatrix}(I^{(r)}, F_{12}) \\ &= \begin{pmatrix} I \\ F_{21} \end{pmatrix} B_{11}(I, F_{12}), \end{aligned} \tag{2.84}$$

其中 F_{12} 与 F_{21} 如 (2.83) 所示. 利用广义逆的运算法则 (见第一章 § 6, 6.2 之 8)),立即得出 (2.82). □

定理 2.13.[157] 设 $A\in \mathbf{C}_r^{m\times n}$, $B=A+E$ 是 A 的锐角扰动.

令

$$\bar{\kappa}' = \|A\| \|B_{11}^{-1}\|_2. \tag{2.85}$$

则

$$\frac{\|B^\dagger - A^\dagger\|}{\|A^\dagger\|} \leqslant \bar{\kappa}' \frac{\|E_{11}\|_2}{\|A\|} + \phi_\Phi\left(\bar{\kappa}' \frac{E_{12}}{\|A\|}\right) + \phi_\Phi\left(\bar{\kappa}' \frac{E_{21}}{\|A\|}\right). \tag{2.86}$$

其中 ϕ_Φ 如 (2.78) 式定义.

证明:

令

$$I_{21} = \begin{pmatrix} I^{(r)} \\ 0 \end{pmatrix}, \quad I_{12} = (I^{(r)}, 0),$$

$$J_{21} = \begin{pmatrix} I^{(r)} \\ F_{21} \end{pmatrix}, \quad J_{12} = (I^{(r)}, F_{12}),$$

其中 $\boldsymbol{F}_{12}$ 与 $\boldsymbol{F}_{21}$ 见 (2.83) 式.

由 (2.82) 可知，$B^\dagger = J_{12}^\dagger B_{11}^{-1} J_{21}^\dagger$. 因此

$$B^\dagger - A^\dagger = (J_{12}^\dagger - I_{12}^\dagger) A_{11}^{-1} I_{21}^\dagger + J_{12}^\dagger A_{11}^{-1}(J_{21}^\dagger - I_{21}^\dagger) + J_{12}^\dagger (B_{11}^{-1} - A_{11}^{-1}) J_{21}^\dagger. \tag{2.87}$$

根据定理 1.1,

$$\|J_{12}^\dagger (B_{11}^{-1} - A_{11}^{-1}) J_{21}^\dagger\| \leqslant \|A_{11}^{-1}\| \bar{\kappa}' \frac{\|E_{11}\|_2}{\|A_{11}\|}. \tag{2.88}$$

根据引理 2.1.

$$\begin{aligned}
\|(J_{12}^\dagger - I_{12}^\dagger) A_{11}^{-1} I_{21}^\dagger\| &\leqslant \|A_{11}^{-1}\| \|J_{12}^\dagger - I_{12}^\dagger\| \\
&= \|A_{11}^{-1}\| \left\| \begin{pmatrix} F_{12} F_{12}^H \\ F_{12}^H \end{pmatrix} (I + F_{12} F_{12}^H)^{-1} \right\| \\
&= \|A_{11}^{-1}\| \phi_\Phi(F_{12}) \\
&= \|A_{11}^{-1}\| \phi_\Phi(B_{11}^{-1} E_{12}) \\
&= \|A_{11}^{-1}\| \phi_\Phi\left(\|A\| B_{11}^{-1} \frac{E_{12}}{\|A\|}\right) \\
&\leqslant \|A_{11}^{-1}\| \phi_\Phi\left(\bar{\kappa}' \frac{E_{12}}{\|A\|}\right),
\end{aligned} \tag{2.89}$$

同理可得

$$\|J_{12}A_{11}^{-1}(J_{21}^{\dagger}-I_{21}^{\dagger})\|\leqslant\|A_{11}^{-1}\|\phi_{\Phi}\left(\bar{\kappa}'\frac{E_{21}}{\|A\|}\right). \tag{2.90}$$

将 (2.87)—(2.90) 联系起来，并利用 $\|A^{\dagger}\|=\|A_{11}^{-1}\|$，即可导出估计式 (2.86). □

注 2.4. 当 $\|E\|$ 足够小时，利用函数 ϕ_{Φ} 的性质 (iii)，从 (2.86) 可得估计式

$$\frac{\|B^{\dagger}-A^{\dagger}\|}{\|A^{\dagger}\|}\lesssim\bar{\kappa}'\frac{\|E_{11}\|_2+\|E_{12}\|+\|E_{21}\|}{\|A\|}. \tag{2.91}$$

此外，当 $m=n=r$ 时，有 $\bar{\kappa}'=\bar{\kappa}$（见(1.9)式)，并且估计式(2.86)与定理 1.1 的估计式(1.8)相吻合. 因此可以说，定理 2.13 是定理 1.1 的推广.

再进一步，如果 $\|E_{11}\|$ 足够小，则可得到定理 1.2 的下述推广.

定理 2.14.[157] 设 $A\in\mathbf{C}_r^{m\times n}$，$B=A+E$ 是 A 的锐角扰动. 令 $\kappa_{\dagger}=\|A\|\|A^{\dagger}\|_2$（同 (2.77)). 如果 E_{11} 满足

$$\|A^{\dagger}\|_2\|E_{11}\|_2<1 \tag{2.92}$$

和

$$\gamma_{\dagger}'\equiv1-\kappa_{\dagger}\|E_{11}\|_2/\|A\|>0, \tag{2.93}$$

则有

$$\|B^{\dagger}\|\leqslant\|A^{\dagger}\|/\gamma_{\dagger} \tag{2.94}$$

和

$$\frac{\|B^{\dagger}-A^{\dagger}\|}{\|A^{\dagger}\|}\leqslant\frac{\kappa_{\dagger}}{\gamma_{\dagger}'}\frac{\|E_{11}\|_2}{\|A\|}+\phi_{\Phi}\left(\frac{\kappa_{\dagger}}{\gamma_{\dagger}'}\frac{E_{12}}{\|A\|}\right)+\phi_{\Phi}\left(\frac{\kappa_{\dagger}}{\gamma_{\dagger}'}\frac{E_{21}}{\|A\|}\right). \tag{2.95}$$

证明:

由 $B^{\dagger}=J_{12}^{\dagger}B_{11}^{-1}J_{21}^{\dagger}$ 可得

$$\|B^{\dagger}\|\leqslant\|J_{12}^{\dagger}\|\|B_{11}^{-1}\|\|J_{21}^{\dagger}\|\leqslant\|B_{11}^{-1}\|. \tag{2.96}$$

根据定理 1.2，

$$\|B_{11}^{-1}\| \leqslant \|A_{11}^{-1}\|/\gamma'_{\dagger} = \|A^{\dagger}\|/\gamma'_{\dagger}. \tag{2.97}$$

代入(2.96)，便得出不等式(2.94).

利用(2.97)可知(2.85)式所示的$\bar{\kappa}'$适合

$$\bar{\kappa}' \leqslant \frac{\|A\|\|A^{\dagger}\|_2}{\gamma'_{\dagger}} = \frac{\kappa_{\dagger}}{\gamma'_{\dagger}}, \tag{2.98}$$

代入(2.86)式的右端，立即导出估计式(2.95). □

注 2.5. 当$m = n = r$时，$\kappa_{\dagger} = \kappa$ (见(1.14)式)，$\gamma'_{\dagger} = \gamma$(见(1.15)式)，条件(2.93)与定理1.2中的条件(1.11)相吻合，并且估计式(2.94)和(2.95)分别与定理1.2中的估计式(1.12)和(1.13)相吻合. 因此可以说，定理2.14是定理1.2的推广.

习题

1. 设 $A, B \in \mathbf{C}^{m\times n}$. 证明恒等式(2.51)和(2.53).

2. 设 $f(\theta) = (\alpha_1\cos\theta + \alpha_2\sin\theta)^2 + \alpha_3^2\cos^2\theta$，其中 $\alpha_1, \alpha_2, \alpha_3 \geqslant 0$. 试证

$$\max_{0\leqslant\theta\leqslant\frac{\pi}{2}} f(\theta) = \frac{1}{2}\left(\alpha_1^2 + \alpha_2^2 + \alpha_3^2 + \sqrt{(\alpha_1^2 + \alpha_2^2 + \alpha_3^2) - 4\alpha_2^2\alpha_3^2}\right).$$

3. 设 $A, B \in \mathbf{C}^{m\times n}$. 如果 $\operatorname{rank}(B) > \operatorname{rank}(A)$，则投影 P_B 必可表示成两个投影之和: $P_B = P_1 + P_2$，其中 $\operatorname{rank}(P_1) = \operatorname{rank}(A)$，并且 $P_A P_2 = 0$.

4. 利用上题的结论证明: 如果 $A, B \in \mathbf{C}^{m\times n}$，满足 $\operatorname{rank}(B) \geqslant \operatorname{rank}(A)$，则对于任一酉不变范数$\|\ \|$，有

$$\|P_B P_A^{\perp}\| \geqslant \|P_B^{\perp} P_A\|.$$

§3 投影的扰动

3.1 关于投影的连续性

例 3.1. $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \\ \varepsilon & 0 \end{pmatrix}$, $\varepsilon > 0$.

显然 $B \to A\ (\varepsilon \to 0)$； 但据定理 2.3 之 2)，$\|P_B - P_A\|_2 = 1 \not\to 0(\varepsilon \to 0)$，即 $P_B \not\to P_A(\varepsilon \to 0)$.

例 3.1 说明，如同广义逆一样，投影一般是不连续的. 下述定理给出了投影连续的必要与充分条件，它与定理 2.12 相类似.

定理 3.1.[23] 设 $A, B \in \mathbb{C}^{m\times n}$，其中 B 视为变量. 则

$$\lim_{B\to A} P_B = P_A$$

的充分与必要条件是当 B 充分靠近 A、并且趋向 A 时，保持

$$\text{rank}(B) = \text{rank}(A).$$

证明:

必要性. 设 $P_B \to P_A(B \to A)$. 如果当 B 充分靠近 A、并且趋向 A 时，不保持 $\text{rank}(B) = \text{rank}(A)$，则存在矩阵序列 $\{B_k\}$: $B_k \to A\ (k \to \infty)$，$\text{rank}(B_k) \neq \text{rank}(A)$，$k = 1, 2, \cdots$. 根据定理 2.3 之 2)，$\|P_{B_k} - P_A\|_2 = 1 \not\to 0(k \to \infty)$，这与 $P_B \to P_A$ $(B \to A)$ 相矛盾.

充分性. 利用恒等式

$$P_B - P_A = (B - A)(B^\dagger - A^\dagger) + A(B^\dagger - A^\dagger) + (B - A)A^\dagger,$$

可得

$$\|P_B - P_A\|_2 \leqslant \|B - A\|_2\|B^\dagger - A^\dagger\|_2 + \|A\|_2\|B^\dagger - A^\dagger\|_2 + \|B - A\|_2\|A^\dagger\|_2. \tag{3.1}$$

如果当 B 充分靠近 A、并且趋向 A 时，保持 $\text{rank}(B) = \text{rank}(A)$，则据定理 2.12，有 $\|B^\dagger - A^\dagger\|_2 \to 0(B \to A)$. 因此，由不等式 (3.1) 立即得出 $P_B \to P_A(B \to A)$. □

注 3.1. Stewart [157] 曾证明了 $P_B \to P_A(B \to A)$ 的一个充分条件是 A 与 B 成锐角，并且 $P_A^\perp B P_{A^H} \to 0(B \to A)$. 定理 3.1 表明，$A$ 与 B 成锐角不必再作为条件列出；此外，由

$$P_A^\perp B P_{A^H} = P_A^\perp(B - A)P_{A^H}$$

可知，当 $B \to A$ 时，必有 $P_A^\perp B P_{A^H} \to 0$，因此，$P_A^\perp B P_{A^H} \to 0$ 也不必单独列为条件.

3.2 投影的扰动界限

首先指出，由定理 2.3 之 2）知，如果

$$A,\ B \in \mathbb{C}^{m\times n},\quad \operatorname{rank}(B) \neq \operatorname{rank}(A),$$

则对于任一酉不变范数$\|\ \|$,有

$$\|P_B - P_A\| \geqslant \|P_B - P_A\|_2 = 1.$$

尽管如此,仍可以求出 $\|P_B - P_A\|$ 的上界.

下面将分别三种不同的情形,进行讨论.

3.2.1. 一般情形

定理 3.2.[23] 设 $A \in \mathbb{C}^{m\times n}$，$B = A + E$. 则有

$$\|P_B - P_A\| \leqslant (\|A^\dagger\|_2 + \|B^\dagger\|_2)\|E\|, \tag{3.2}$$

$$\|P_B - P_A\|_F \leqslant \sqrt{\|A^\dagger\|_2^2 + \|B^\dagger\|_2^2}\|E\|_F \tag{3.3}$$

和

$$\|P_B - P_A\|_2 \leqslant \max\{\|A^\dagger\|_2, \|B^\dagger\|_2\}\|E\|_2. \tag{3.4}$$

证明:

1）利用

$$\begin{aligned}
P_B - P_A &= P_B(I - P_A) - (I - P_B)P_A \\
&= B^{\dagger H}B^H(I - A^{\dagger H}A^H) - (I - BB^\dagger)AA^\dagger \\
&= B^{\dagger H}(B - A)^H(I - A^{\dagger H}A^H) \\
&\quad + (I - BB^\dagger)(B - A)A^\dagger \\
&= B^{\dagger H}E^H P_A^\perp + P_B^\perp E A^\dagger,
\end{aligned} \tag{3.5}$$

立即得出不等式 (3.2).

2）利用 (3.5) 式可知

$$\begin{aligned}
\|P_B - P_A\|_F^2 &= \operatorname{tr}((P_B - P_A)^2) \\
&= \|B^{\dagger H}E^H P_A^\perp\|_F^2 + \|P_B^\perp E A^\dagger\|_F^2 \\
&\leqslant (\|B^\dagger\|_2^2 + \|A^\dagger\|_2^2)\|E\|_F^2,
\end{aligned}$$

即 (3.3) 式成立.

3）将 (3.5) 式改写成

$$P_B - P_A = P_B B^{\dagger H}E^H P_A^\perp + P_B^\perp E A^\dagger P_A. \tag{3.6}$$

任取 $\boldsymbol{u} \in \mathbb{C}^m$，$\|\boldsymbol{u}\|_2 = 1$,便有

$$\|(P_B - P_A)\boldsymbol{u}\|_2^2 \leqslant \|P_B B^{\dagger H} E^H\|_2^2 \|P_A^{\perp}\boldsymbol{u}\|_2^2 + \|P_B^{\perp} E A^{\dagger}\|_2^2 \|P_A \boldsymbol{u}\|_2^2. \quad (3.7)$$

令

$$\|P_B B^{\dagger} E^H\|_2 = \alpha \geqslant 0, \quad \|P_B^{\perp} E A^{\dagger}\|_2 = \beta$$

和

$$\|P_A^{\perp}\boldsymbol{u}\|_2 = \cos\theta, \quad \|P_A \boldsymbol{u}\|_2 = \sin\theta, \quad 0 \leqslant \theta \leqslant \frac{\pi}{2}.$$

则不等式 (3.7) 可写为

$$\|(P_B - P_A)\boldsymbol{u}\|_2^2 \leqslant \alpha^2\cos^2\theta + \beta^2\sin^2\theta \leqslant \max\{\alpha^2, \beta^2\} \leqslant \max\{\|A^{\dagger}\|_2^2, \|B^{\dagger}\|_2^2\}\|E\|_2^2,$$

因此不等式 (3.4) 成立. □

注意，当 $B \to A$ 时，虽有 $\|E\| \to 0$，但 $\|A^{\dagger}\|_2$ 或 $\|B^{\dagger}\|_2$ 可能是无界的.

3.2.2. $\mathrm{rank}(A) = \mathrm{rank}(B)$ 的情形

定理 3.3.[23] 设 $A \in \mathbf{C}^{m\times n}$，$B = A + E$. 如果 $\mathrm{rank}(A) = \mathrm{rank}(B)$，则有

$$\|P_B - P_A\| \leqslant 2\min\{\|A^{\dagger}\|_2, \|B^{\dagger}\|_2\}\|E\|, \quad (3.8)$$

$$\|P_B - P_A\|_F \leqslant \sqrt{2}\min\{\|A^{\dagger}\|_2, \|B^{\dagger}\|_2\}\|E\|_F \quad (3.9)$$

和

$$\|P_B - P_A\|_2 \leqslant \min\{\|A^{\dagger}\|_2, \|B^{\dagger}\|_2\}\|E\|_2. \quad (3.10)$$

证明:

根据 (2.36) 和 (2.37)，有

$$\sigma(P_A P_B^{\perp}) = \{\sigma_1, \sigma_2, \cdots, \sigma_{r_1}, 0, \cdots, 0\},$$

$$\sigma(P_B - P_A) = \{\sigma_1, \sigma_1, \cdots, \sigma_{r_1}, \sigma_{r_1}, 0, \cdots, 0\}.$$

由此可得

$$\|P_B - P_A\|_2 = \|P_A P_B^{\perp}\|_2 = \|P_B P_A^{\perp}\|_2 \quad (\text{根据 } (2.33))$$
$$\leqslant \min\{\|A^{\dagger}\|_2, \|B^{\dagger}\|_2\}\|E\|_2 \quad (\text{根据 } (2.43))$$

和

$$\|P_B - P_A\|_F = \sqrt{2}\|P_A P_B^{\perp}\|_F = \sqrt{2}\|P_B P_A^{\perp}\|_F \ (\text{根据 } (2.33))$$
$$\leqslant \sqrt{2}\min\{\|A^{\dagger}\|_2, \|B^{\dagger}\|_2\}\|E\|_F \ (\text{根据 } (2.43)).$$

此外，容易看出，对每一个特殊的酉不变范数$\|\ \ \|_{(k)}$（见第二章§3例3.3），有

$$\|P_B - P_A\|_{(k)} \leqslant 2\|P_A P_B^{\perp}\|_{(k)}, \quad k = 1, \cdots, m,$$

因此，根据第二章定理3.6，对于 $\mathbb{C}^{m\times m}$ 上的任一酉不变范数$\|\ \ \|$，有

$$\begin{aligned}\|P_B - P_A\| &\leqslant 2\|P_A P_B^{\perp}\| = 2\|P_B P_A^{\perp}\| \quad (\text{根据}(2.33))\\ &\leqslant 2\min\{\|A^{\dagger}\|_2, \|B^{\dagger}\|_2\}\|E\| \quad (\text{根据}(2.43)). \ \square\end{aligned}$$

当B靠近A时，可以使用估计式(3.8)—(3.10)，给出$\|P_B - P_A\|$的上界．但是，当B并不靠近A时，$\|P_B - P_A\|$也可能很小；如果在这时仍然使用估计式(3.8)—(3.10)，则不可能得到令人满意的结果，试看下例：

例 3.2. $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ 0 & \varepsilon \\ 0 & 0 \end{pmatrix}$, $0 < \varepsilon \ll 1$.

容易算出

$$A^{\dagger} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad B^{\dagger} = \frac{1}{1+\varepsilon^2}\begin{pmatrix} 0 & 0 & 0 \\ 1 & \varepsilon & 0 \end{pmatrix},$$

$$E = \begin{pmatrix} -1 & 1 \\ 0 & \varepsilon \\ 0 & 0 \end{pmatrix}, \quad P_A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$P_B = \frac{1}{1+\varepsilon^2}\begin{pmatrix} 1 & \varepsilon & 0 \\ \varepsilon & \varepsilon^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

和

$$\|P_B - P_A\|_2 = \frac{\varepsilon}{\sqrt{1+\varepsilon^2}} \approx \varepsilon;$$

但(3.10)式的右端为

$$\begin{aligned}\min\left\{1, \frac{1}{\sqrt{1+\varepsilon^2}}\right\}\left\|\begin{pmatrix} -1 & 1 \\ 0 & \varepsilon \end{pmatrix}\right\|_2 &= \sqrt{\frac{2+\varepsilon^2+\sqrt{4+\varepsilon^4}}{2(1+\varepsilon^2)}}\\ &\approx 1.\end{aligned}$$

因此，为了得到 $\|P_B - P_A\|$ 的比较小的上界，可以利用 A 与 B 的满秩分解

$$A = A_1F_1, \quad B = B_1G_1, \tag{3.11}$$

其中 $A_1, B_1 \in \mathbb{C}_r^{m\times r}$，$F_1, G_1 \in \mathbb{C}_r^{r\times n}$. 显然有

$$P_A = P_{A_1}, \quad P_B = P_{B_1}. \tag{3.12}$$

所以从定理 3.3 立即得到

定理 3.4.[23] 设 $A \in \mathbb{C}^{m\times n}, B = A + E, \operatorname{rank}(A) = \operatorname{rank}(B)$. 任取 A 与 B 的满秩分解 (3.11)，则有

$$\|P_B - P_A\| \leqslant \mu \min\{\|A_1^\dagger\|_2, \|B_1^\dagger\|_2\}\|B_1 - A_1\| \tag{3.13}$$

$$= \mu\|B_1 - A_1\|/\max\{\sigma_{\min}(A_1), \sigma_{\min}(B_1)\}, \tag{3.14}$$

其中

$$\mu = \begin{cases} 2 & \text{对于任意的酉不变范数} \|\ \ \| \\ \sqrt{2} & \text{对于} \|\ \ \|_F \\ 1 & \text{对于} \|\ \ \|_2, \end{cases} \tag{3.15}$$

$\sigma_{\min}(C)$ 表示矩阵 C 的最小奇异值.

对于例 3.2 中的 A 与 B，可以分别取

$$A_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \text{与 } B = \begin{pmatrix} 1 \\ \varepsilon \\ 0 \end{pmatrix}.$$

根据 (3.13)—(3.15)，可得估计式

$$\|P_B - P_A\| \leqslant \mu \min\left\{1, \frac{1}{\sqrt{1+\varepsilon^2}}\right\} \left\|\begin{pmatrix} 0 \\ \varepsilon \\ 0 \end{pmatrix}\right\| = \mu\varepsilon / \sqrt{1+\varepsilon^2}.$$

其中 μ 如 (3.15) 式所示.

3.2.3. 锐角扰动的情形

利用 A 与 B 的约简形式(2.8)与(2.9)，Stewart 证明了下述结论.

定理 3.5.[157] 设 $A \in \mathbb{C}_r^{m\times n}$, $B = A + E$ 是 A 的锐角扰动. 则

$$\|P_B - P_A\|_2 \leqslant \frac{\bar{\kappa}'\|E_{21}\|_2/\|A\|}{[1 + (\bar{\kappa}'\|E_{21}\|_2/\|A\|)^2]^{1/2}} < 1. \tag{3.16}$$

其中 $\tilde{\kappa}'$ 如(2.85)式所示.

证明:

根据 B 是 A 的锐角扰动,可知

$$R(B)=R\left(\begin{pmatrix} I \\ F_{21} \end{pmatrix}\right),\quad F_{21}=E_{21}B_{11}^{-1}.$$

所以

$$P_B=P_{\binom{I}{F_{21}}}=\begin{pmatrix} I \\ F_{21} \end{pmatrix}(I+F_{21}^HF_{21})^{-1}(I,F_{21}).$$

由此可得

$$P_B-P_A=\begin{pmatrix} (I+F_{21}^HF_{21})^{-1}-I & (I+F_{21}^HF_{21})^{-1}F_{21}^H \\ F_{21}(I+F_{21}^HF_{21})^{-1} & F_{21}(I+F_{21}^HF_{21})^{-1}F_{21}^H \end{pmatrix}$$

和

$$(P_B-P_A)^2=\begin{pmatrix} F_{21}^HF_{21}(I+F_{21}^HF_{21})^{-1} & 0 \\ 0 & F_{21}(I+F_{21}^HF_{21})^{-1}F_{21}^H \end{pmatrix}. \tag{3.17}$$

设 $\{\sigma_i(F_{21})\}_{i=1}^{r_1}$ 是 F_{21} 的非零奇异值, $r_1\leqslant r$; 则由(3.17)可知 $(P_B-P_A)^2$ 的非零奇异值为

$$\frac{\sigma_i^2(F_{21})}{1+\sigma_i^2(F_{21})},\quad i=1,\cdots,r_1.$$

再注意到 F_{21} 的最大奇异值

$$\sigma_1(F_{21})=\|F_{21}\|_2\leqslant\|E_{21}\|_2\|B_{11}^{-1}\|_2$$

$$=\|A\|\|B_{11}^{-1}\|_2\frac{\|E_{21}\|_2}{\|A\|}=\tilde{\kappa}'\frac{\|E_{21}\|_2}{\|A\|},$$

代入

$$\|P_B-P_A\|_2=\frac{\sigma_1(F_{21})}{\sqrt{1+\sigma_1^2(F_{21})}}$$

便导出了不等式(3.16). □

注 3.2. 估计式(3.16)可以写成一般的形式:

$$\|P_B-P_A\|_2\leqslant\frac{\tilde{\kappa}'\|P_A^{\perp}EP_{A^H}\|_2/\|A\|}{[1+(\tilde{\kappa}'\|P_A^{\perp}EP_{A^H}\|_2/\|A\|)^2]^{1/2}}<1, \tag{3.18}$$

其中

$$\tilde{\kappa}' = \|A\|\|(P_A B P_{A^H})^{\dagger}\|_2. \tag{3.19}$$

习题

1. 设 $A\in\mathbb{C}^{m\times n}$, $B=A+E$. 证明恒等式

$$P_B - P_A = B^{\dagger H}P_{B^H}E^H P_A^{\perp} + P_B^{\perp}E P_{A^H}A^{\dagger}.$$

2. 设 $A\in\mathbb{C}^{m\times n}$, $B=A+E$ 满足 $\text{rank}(B)=\text{rank}(A)$. 试证渐近公式

$$P_B = P_A + P_A^{\perp}E P_{A^H}A^{\dagger} + A^{\dagger H}P_{A^H}E^H P_A^{\perp} + O(\|E\|^2).$$

§4 线性最小二乘问题扰动分析

设 $A\in\mathbb{C}^{m\times n}$, $\boldsymbol{b}\in\mathbb{C}^m$. 第一章§7定理7.5已经指出：线性最小二乘问题

$$\|A\boldsymbol{x}-\boldsymbol{b}\| = \min \tag{4.1}$$

的最小范数解(即同时满足 $\|\boldsymbol{x}\|=\min$ 的解)为 $\boldsymbol{x}=A^{\dagger}\boldsymbol{b}$. 本节将讨论 A 与 $\boldsymbol{b}$ 的扰动对于问题(4.1)的最小范数解 $\boldsymbol{x}$ 的影响.

设 $B=A+E$, $\boldsymbol{c}=\boldsymbol{b}+\boldsymbol{k}$, 线性最小二乘问题

$$\|B\boldsymbol{x}-\boldsymbol{c}\| = \min \tag{4.2}$$

的最小范数解为 $\boldsymbol{x}+\boldsymbol{h}$.

首先讨论 $\text{rank}(A)=\text{rank}(B)$ 的情形.

定理 4.1.[157] 设 $A\in\mathbb{C}^{m\times n}$, $B=A+E$, $\boldsymbol{c}=\boldsymbol{b}+\boldsymbol{k}\in\mathbb{C}^m$. 又设线性最小二乘问题(4.1)和(4.2)的解分别为 $\boldsymbol{x}$ 和 $\boldsymbol{x}+\boldsymbol{h}$. 如果 $\text{rank}(A)=\text{rank}(B)$, 并且 $\|A^{\dagger}\|_2\|E\|_2<1$, 则

$$\|\boldsymbol{h}\| \leqslant \frac{\kappa_{\dagger}}{\gamma_{\dagger}}\left(\frac{\|E\|_2}{\|A\|}\|\boldsymbol{x}\| + \frac{\|\boldsymbol{k}\|}{\|A\|} + \frac{\kappa_{\dagger}}{\gamma_{\dagger}}\frac{\|E\|_2}{\|A\|}\frac{\|\boldsymbol{r}_x\|}{\|A\|} + \|E\|_2\|\boldsymbol{y}\|\right), \tag{4.3}$$

其中 $\kappa_{\dagger}$ 和 $\gamma_{\dagger}$ 如(2.77)式所示，并且

$$\boldsymbol{r}_x = \boldsymbol{b} - A\boldsymbol{x} \tag{4.4}$$

和

$$y = A^{\dagger H} x; \tag{4.5}$$

特别地，当 $\operatorname{rank}(A) = \operatorname{rank}(B) = n$，并且 $\|A^{\dagger}\|_2 \|E\|_2 < 1$ 时，有

$$\|h\| \leqslant \frac{\kappa_{\dagger}}{\gamma_{\dagger}} \left(\frac{\|E\|_2}{\|A\|} \|x\| + \frac{\|k\|}{\|A\|} + \frac{\kappa_{\dagger}}{\gamma_{\dagger}} \frac{\|E\|_2}{\|A\|} \frac{\|r_x\|}{\|A\|} \right). \tag{4.6}$$

证明:

由 $x = A^{\dagger} b$ 和 $x + h = B^{\dagger}(b + k)$ 可知

$$\begin{aligned} h &= B^{\dagger}(b + k) - A^{\dagger} b = (B^{\dagger} - A^{\dagger}) b + B^{\dagger} k \\ &= (-B^{\dagger} E A^{\dagger} + B^{\dagger} P_A^{\perp} - P_{B^H}^{\perp} A^{\dagger}) b + B^{\dagger} k \quad (\text{根据 } (2.51)) \\ &= (-B^{\dagger} E x + B^{\dagger} r_x + B^{\dagger} k) - P_{B^H}^{\perp} x. \end{aligned} \tag{4.7}$$

利用不等式 (2.75)，易得估计式

$$\begin{aligned} \|B^{\dagger} E x\| &\leqslant \|B^{\dagger}\|_2 \|E\|_2 \|x\| \\ &\leqslant \frac{\|A^{\dagger}\|_2}{\gamma_{\dagger}} \|E\|_2 \|x\| \\ &= \frac{\kappa_{\dagger}}{\gamma_{\dagger}} \frac{\|E\|_2}{\|A\|} \|x\|, \end{aligned} \tag{4.8}$$

$$\begin{aligned} \|B^{\dagger} r_x\| &= \|B^{\dagger} P_B P_A^{\perp} r_x\| \leqslant \|B^{\dagger}\| \|P_B P_A^{\perp}\|_2 \|r_x\| \\ &= \|B^{\dagger}\|_2 \|B^{\dagger H}(B - A)^H (I - A^{\dagger H} A^H)\|_2 \|r_x\| \\ &\leqslant \|B^{\dagger}\|_2^2 \|E\|_2 \|r_x\| \leqslant \frac{\|A^{\dagger}\|_2^2}{\gamma_{\dagger}^2} \|E\|_2 \|r_x\| \\ &= \frac{\kappa_{\dagger}^2}{\gamma_{\dagger}^2} \frac{\|E\|_2}{\|A\|} \frac{\|r_x\|}{\|A\|}, \end{aligned} \tag{4.9}$$

$$\|B^{\dagger} k\| \leqslant \|B^{\dagger}\|_2 \|k\| \leqslant \frac{\kappa_{\dagger}}{\gamma_{\dagger}} \frac{\|k\|}{\|A\|}. \tag{4.10}$$

和

$$\begin{aligned} \|P_{B^H}^{\perp} x\| &= \|(I - B^H B^{H\dagger}) A^{\dagger} A A^{\dagger} b\| \\ &= \|(I - B^H B^{H\dagger}) A^H A^{\dagger H} x\| \\ &= \|(I - B^H B^{H\dagger})(B - A)^H A^{\dagger H} x\| \\ &\leqslant \|E\|_2 \|A^{\dagger H} x\| = \|E\|_2 \|y\|. \end{aligned} \tag{4.11}$$

将 (4.7)—(4.11) 联系起来，便得出估计式 (4.3)。

再注意到，当 $\text{rank}(A)=\text{rank}(B)=n$ 时，有 $P_{B^H}^{\perp}=0$，因而 (4.7) 式右端最后一项为零. 所以有不等式 (4.6). □

注 4.1. 当 $A\in\mathbb{C}^{n\times n}, B=A+E, \text{rank}(A)=\text{rank}(B)=n$ 时，由(4.6)式可知，方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 的解 $\boldsymbol{x}$ 与方程组 $B(\boldsymbol{x}+\boldsymbol{h})=\boldsymbol{b}+\boldsymbol{k}$ 的解 $\boldsymbol{x}+\boldsymbol{h}$，满足

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|}\leqslant\frac{\kappa}{\gamma}\left(\frac{\|E\|_2}{\|A\|}+\frac{\|\boldsymbol{k}\|}{\|A\|\|\boldsymbol{x}\|}\right)$$

$$\leqslant\frac{\kappa}{\gamma}\left(\frac{\|E\|_2}{\|A\|}+\frac{\|\boldsymbol{k}\|}{\|\boldsymbol{b}\|}\right).$$

这与定理 1.3 中的估计式 (1.20) 相吻合.

当 B 是 A 的锐角扰动时，根据分解式(2.6)与 (2.7)，可以分别把 $V^H\boldsymbol{x}$ 和 $U^H\boldsymbol{b}$ 记作 $\boldsymbol{x}$ 和 $\boldsymbol{b}$，因此无妨假定 A 与 B 具有约简形式 (2.8)与(2.9). 这时，将 $\boldsymbol{x}$ 与 $\boldsymbol{b}$ 进行相应的分块:

$$\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{pmatrix}\begin{matrix}r\\ n-r\end{matrix},\quad \boldsymbol{b}=\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix}\begin{matrix}r\\ m-r\end{matrix}. \tag{4.12}$$

于是，最小二乘问题 $\|A\boldsymbol{x}-\boldsymbol{b}\|=\min$ 的最小范数解 $\boldsymbol{x}$ 为

$$\boldsymbol{x}_1=A_{11}^{-1}\boldsymbol{b}_1,\quad \boldsymbol{x}_2=0. \tag{4.13}$$

此外，剩余向量

$$\boldsymbol{r}_x=\boldsymbol{b}-A\boldsymbol{x}$$

适合

$$\|\boldsymbol{r}_x\|=\|\boldsymbol{b}_2\|.$$

在下面的讨论中，常数 $\kappa_\dagger$ 与 $\bar{\kappa}'$ 分别如(2.77)式和(2.85)式所示. 另外，定义 η 为满足

$$\|\boldsymbol{b}_1\|=\eta\|A\|\|\boldsymbol{x}\| \tag{4.14}$$

的非负常数. 由 $\boldsymbol{b}_1=A_{11}\boldsymbol{x}$ 可推知 $\eta\leqslant 1$；再由 $\|\boldsymbol{x}\|\leqslant\|A^\dagger\|_2\|\boldsymbol{b}_1\|$ 可推知 $\eta\geqslant\kappa_\dagger^{-1}$. 所以 η 满足

$$\kappa_\dagger^{-1}\leqslant\eta\leqslant 1. \tag{4.15}$$

定理 4.2.[157] 设 $\boldsymbol{x}=A^\dagger\boldsymbol{b}$，$\boldsymbol{x}+\boldsymbol{h}=A^\dagger(\boldsymbol{b}+\boldsymbol{k})$. 则

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|}\leqslant\kappa_\dagger\eta\|P_A\boldsymbol{k}\|/\|P_A\boldsymbol{b}\|. \tag{4.16}$$

证明:

利用分块 (4.12) 和

$$\boldsymbol{h}=\begin{pmatrix}\boldsymbol{h}_1\\ \boldsymbol{h}_2\end{pmatrix}\begin{matrix}r\\ n-r\end{matrix},\ \boldsymbol{k}=\begin{pmatrix}\boldsymbol{k}_1\\ \boldsymbol{k}_2\end{pmatrix}\begin{matrix}r\\ m-r\end{matrix},\tag{4.17}$$

有

$$\boldsymbol{h}_1=A_{11}^{-1}\boldsymbol{k}_1,\quad \boldsymbol{h}_2=0.$$

因此

$$\|\boldsymbol{h}\|\leqslant\|A_{11}^{-1}\|_2\|\boldsymbol{k}_1\|.\tag{4.18}$$

根据 (4.14)，$\|\boldsymbol{x}\|=\|\boldsymbol{b}_1\|/\eta\|A\|$. 与 (4.18) 联系起来，便得到

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|}\leqslant\frac{\eta\|A\|\|A_{11}^{-1}\|_2\|\boldsymbol{k}_1\|}{\|\boldsymbol{b}_1\|}=\kappa_{\dagger}\eta\frac{\|P_A\boldsymbol{k}\|}{\|P_A\boldsymbol{b}\|}.\quad\square$$

注 4.2. 设 $\boldsymbol{x}$ 是最小二乘问题(4.1)的最小范数解. 利用分块(4.12)和关系式(4.13)，可知 $\boldsymbol{b}=A\boldsymbol{x}+\boldsymbol{r}_x$ 具有下述形式:

$$\begin{pmatrix}\boldsymbol{b}_1\\ \boldsymbol{b}_2\end{pmatrix}=\begin{pmatrix}A_{11}&0\\ 0&0\end{pmatrix}\begin{pmatrix}\boldsymbol{x}_1\\ 0\end{pmatrix}+\begin{pmatrix}0\\ \boldsymbol{r}_x^{(2)}\end{pmatrix}.\tag{4.19}$$

因此，

$$\|\boldsymbol{b}\|^2=\|\boldsymbol{b}_1\|^2+\|\boldsymbol{r}_x^{(2)}\|^2=\|P_A\boldsymbol{b}\|^2+\|\boldsymbol{r}_x\|^2.$$

代入 (4.16)，得到

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|}\leqslant\kappa_{\dagger}\eta\|P_A\boldsymbol{k}\|/\sqrt{\|\boldsymbol{b}\|^2-\|\boldsymbol{r}_x\|^2}.\tag{4.20}$$

上式表明: 剩余量 $\|\boldsymbol{r}_x\|$ 愈大，则最小二乘解 $\boldsymbol{x}$ 的扰动愈大.

定理 4.3.[157] 设 $\boldsymbol{x}=A^{\dagger}\boldsymbol{b},\boldsymbol{x}+\boldsymbol{h}=B^{\dagger}\boldsymbol{b}$，其中 B 是 A 的锐角扰动，则

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|}\leqslant\bar{\kappa}'\frac{\|E_{11}\|_2}{\|A\|}+\phi_2\left(\bar{\kappa}'\frac{E_{12}}{\|A\|}\right)+\bar{\kappa}'^2\frac{\|E_{12}\|_2}{\|A\|}\left(\eta\frac{\|\boldsymbol{b}_2\|}{\|\boldsymbol{b}_1\|}+\frac{\|E_{21}\|_2}{\|A\|}\right).\tag{4.21}$$

其中 ϕ_2 是对应于谱范数$\|\ \|_2$的函数 ϕ_Φ (见 (2.78) 式和 ϕ_Φ 的性质 (v)).

证明:

利用 (2.82) 和 (2.83)，以及定理 2.13 证明中的记号，有

$$\boldsymbol{h} = B^{\dagger}\boldsymbol{b} - A^{\dagger}\boldsymbol{b} = J_{12}^{\dagger}(B_{11}^{-1} - A_{11}^{-1})\boldsymbol{b}_1 + (J_{12}^{\dagger} - I_{12}^{\dagger})A_{11}^{-1}\boldsymbol{b}_1 + J_{12}^{\dagger}B_{11}^{-1}(J_{21}^{\dagger} - I_{21}^{\dagger})\boldsymbol{b}. \tag{4.22}$$

对上式右端诸项，分别估计如下：

$$\|J_{12}^{\dagger}(B_{11}^{-1} - A_{11}^{-1})\boldsymbol{b}_1\| \leqslant \bar{\kappa}' \frac{\|E_{11}\|_2}{\|A\|}\|\boldsymbol{x}\|; \tag{4.23}$$

$$\|(J_{12}^{\dagger} - I_{12}^{\dagger})A_{11}^{-1}\boldsymbol{b}_1\| \leqslant \phi_2\left(\bar{\kappa}' \frac{E_{12}}{\|A\|}\right)\|\boldsymbol{x}\|; \tag{4.24}$$

最后一项

$$J_{12}^{\dagger}B_{11}^{-1}(J_{21}^{\dagger} - I_{21}^{\dagger})\boldsymbol{b} = J_{12}^{\dagger}B_{11}^{-1}[(I + F_{21}^{H}F_{21})^{-1} - I]\boldsymbol{b}_1 + J_{12}^{\dagger}B_{11}^{-1}(I + F_{21}^{H}F_{21})^{-1}F_{21}^{H}\boldsymbol{b}_2, \tag{4.25}$$

其中

$$\begin{aligned}
&\|J_{12}^{\dagger}B_{11}^{-1}[(I + F_{21}^{H}F_{21})^{-1} - I]\boldsymbol{b}_1\| \\
&= \|J_{12}^{\dagger}B_{11}^{-1}(I + F_{21}^{H}F_{21})^{-1}F_{21}^{H}F_{21}\boldsymbol{b}_1\| \\
&\leqslant \|B_{11}^{-1}\|_2\|(I + F_{21}^{H}F_{21})^{-1}\|_2\|F_{21}^{H}\|_2\|F_{21}\boldsymbol{b}\| \\
&\leqslant \|B_{11}^{-1}\|_2^2\|E_{21}\|_2\|E_{21}B_{11}^{-1}\boldsymbol{b}_1\| \\
&\leqslant \|B_{11}^{-1}\|_2^2\|E_{21}\|_2^2\|\boldsymbol{x}\| \\
&= \left(\bar{\kappa}' \frac{\|E_{21}\|_2}{\|A\|}\right)^2\|\boldsymbol{x}\|,
\end{aligned} \tag{4.26}$$

$$\begin{aligned}
&\|J_{21}^{\dagger}B_{11}^{-1}(I + F_{21}^{H}F_{21})^{-1}F_{21}\boldsymbol{b}_2\| \\
&\leqslant \|B_{11}^{-1}\|_2^2\|E_{21}\|_2\|\boldsymbol{b}_2\| \\
&= \|B_{11}^{-1}\|_2^2\|E_{21}\|_2\frac{\|\boldsymbol{b}_2\|}{\|\boldsymbol{b}_1\|}\eta\|\boldsymbol{x}\|\|A\| \\
&= \eta\bar{\kappa}'^2\frac{\|E_{21}\|_2}{\|A\|}\frac{\|\boldsymbol{b}_2\|}{\|\boldsymbol{b}_1\|}\|\boldsymbol{x}\|.
\end{aligned} \tag{4.27}$$

把关系式(4.22)—(4.27)联系起来，便得到估计式 (4.21)．□

注 4.3. 当 $A \in \mathbf{C}^{n\times n}, B = A + E$, $\mathrm{rank}(A) = \mathrm{rank}(B)$ 时，由(4.21)可知，方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的解 $\boldsymbol{x}$ 与方程组 $B(\boldsymbol{x} + \boldsymbol{h}) = \boldsymbol{b}$ 的解 $\boldsymbol{x} + \boldsymbol{h}$，满足

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|} \leqslant \bar{\kappa}\frac{\|E\|_2}{\|A\|}, \quad \bar{\kappa} = \|A\|\|B^{-1}\|_2.$$

根据定理 1.2，当 $\|A^{-1}\|_2\|E\|_2<1$ 时，有 $\|B^{-1}\|_2\leqslant\|A^{-1}\|_2/\gamma$，从而得到

$$\bar{\kappa}\leqslant\|A\|\|A^{-1}\|_2/\gamma=\kappa/\gamma$$

和

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|}\leqslant\frac{\kappa}{\gamma}\frac{\|E\|_2}{\|A\|}.$$

这与定理 1.3 的结论相吻合.

下面通过一个简单的例子，说明本节的定理.

例 4.1.

$$A=\begin{pmatrix}1 & 0.9999\\ 1 & 1\\ 0 & 0\end{pmatrix},\ \boldsymbol{b}=\begin{pmatrix}1\\ 0\\ 1\end{pmatrix},\ E=\begin{pmatrix}0 & 0.0001\\ 0 & -0.0001\\ 0 & 0\end{pmatrix},$$

$$B=A+E=\begin{pmatrix}1 & 1\\ 1 & 0.9999\\ 0 & 0\end{pmatrix},\ \boldsymbol{c}=\begin{pmatrix}1\\ 0\\ 1\end{pmatrix}.$$

计算可知

$$A^{\dagger}=\begin{pmatrix}10000 & -9999 & 0\\ -10000 & 10000 & 0\end{pmatrix},\ \boldsymbol{x}=A^{\dagger}\boldsymbol{b}=\begin{pmatrix}10000\\ -10000\end{pmatrix},$$

$$B^{\dagger}=\begin{pmatrix}-9999 & 10000 & 0\\ 10000 & -10000 & 0\end{pmatrix},$$

$$\boldsymbol{x}+\boldsymbol{h}=B^{\dagger}\boldsymbol{c}=\begin{pmatrix}-9999\\ 10000\end{pmatrix},\ \boldsymbol{h}=\begin{pmatrix}-19999\\ 20000\end{pmatrix}$$

和

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|}\approx\frac{28284}{14142}=2.$$

注意到 $\operatorname{rank}(A)=\operatorname{rank}(B)=2$，并且 B 是 A 的锐角扰动，所以可以考虑直接应用定理 4.1 或定理 4.3 去估计 $\|\boldsymbol{h}\|/\|\boldsymbol{x}\|$.

利用谱范数，可算出

$$\kappa_{\dagger}=\|A\|_2\|A^{\dagger}\|_2\approx1.99995\times19999.5\approx39998$$

和

$$\gamma_{\dagger} = 1 - \|A^{\dagger}\|_2\|E\|_2 = 1 - 19999.5 \times 0.0001414$$
$$\approx -1.8279;$$
因为 $\gamma_{\dagger} < 0$，所以不能应用定理 4.1.

现考虑应用定理 4.3. 因为 B 中的 $E_{12} = 0$，所以估计式 (4.21) 的右端只有第一项. 于是得到

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{x}\|} \leqslant \bar{\kappa}' \frac{\|E_{11}\|_2}{\|A\|_2} = \|A\|_2\|B_{11}^{-1}\|_2 \cdot \frac{\|E_{11}\|_2}{\|A\|_2}$$
$$= \|B_{11}^{-1}\|_2\|E_{11}\|_2$$
$$\approx 19999 \times 0.0001414 \approx 2.828.$$

估计式 (4.21) 中的 $\bar{\kappa}'$ 是最小二乘问题解的条件数，对于例 4.1，利用谱范数可算出

$$\bar{\kappa}' = \|A\|_2\|B_{11}^{-1}\|_2 \approx 1.99995 \times 19999 \approx 39997 \gg 1;$$

因此，系数矩阵 A 的微小扰动，引起了最小范数解 $\boldsymbol{x}$ 的显著改变.

习题

1. 设 $A \in \mathbb{C}^{m\times n}$，$B = A + E$ 满足 $\mathrm{rank}(B) = \mathrm{rank}(A)$，$\boldsymbol{b} \in \mathbb{C}^m$，$\boldsymbol{c} = \boldsymbol{b} + \boldsymbol{k}$，$\boldsymbol{x}$ 是 $\|A\boldsymbol{x} - \boldsymbol{b}\| = \min$ 的最小范数解，$\tilde{\boldsymbol{x}}$ 是 $\|B\boldsymbol{x} - \boldsymbol{c}\| = \min$ 的最小范数解. 定义剩余 $\boldsymbol{r}_x = \boldsymbol{b} - A\boldsymbol{x}$ 和 $\tilde{\boldsymbol{r}}_x = \boldsymbol{c} - B\boldsymbol{x}$. 试证不等式

$$\|\tilde{\boldsymbol{r}}_x - \boldsymbol{r}_x\| \leqslant \max\{\|A^{\dagger}\|_2, \|B^{\dagger}\|_2\}\|E\|_2\|\boldsymbol{b}\| + \|\boldsymbol{k}\|.$$

2. 设 $A \in \mathbb{C}^{m\times n}$，$B = A + E$ 满足 $\mathrm{rank}(B) = \mathrm{rank}(A)$，$\boldsymbol{b} \in \mathbb{C}^m$，$\boldsymbol{x}$ 与 $\tilde{\boldsymbol{x}}$ 分别是 $\|A\boldsymbol{x} - \boldsymbol{b}\| = \min$ 与 $\|B\boldsymbol{x} - \boldsymbol{b}\| = \min$ 的最小范数解. 证明渐近公式

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} - A^{\dagger}P_AEP_{A^H}\boldsymbol{x} + P_{A^H}^{\perp}E^HP_A^{\dagger}\boldsymbol{x} + (A^HA)^{\dagger}P_{A^H}E^HP_A^{\perp}\boldsymbol{b}$$
$$+ O(\|E\|^2).$$

第五章说明

本章以 Stewart 的综合性论文[157]为线索.

关于线性方程组解的扰动，主要参考[138]和[171]；关于广义逆的扰动，主要参考[53]、[151]和[183]；关于投影的扰动，根据作

者的论文[23];关于线性最小二乘问题解的扰动,主要参考[133]、[157]和[172].

值得指出的是,近几年来,一些数值代数专家正在研究某些特殊类型的最小二乘问题,例如带有线性约束或二次约束的最小二乘问题,以及完全最小二乘问题,等等,它们有重要的应用背景.有关的问题及其扰动分析,读者可参阅[84]、[85]、[93]、[98]和[133].

参 考 文 献

[1] 王国荣、匡蛟勋,关于矩阵病态程度的一种新度量,高等学校计算数学学报，**1**:1 (1979),21—30.

[2] 石钟慈，具有正特征值矩阵的P条件数界限的估计，数学学报，**14**:6 (1964), 790—795.

[3] 石钟慈,王伯英,某些矩阵的行列式,特征值以及条件数界限的若干估计,数学学报，**15**:3(1965)，326—341.

[4] 华罗庚，多复变数函数论中的典型域的调和分析，科学出版社，1958.

[5] 祁力群，Some simple estimates for singular values of a matrix *Lin. Alg.* & *Its Applic.*, **56**(1984), 105—119.

[6] 孙继广,不变子空间与广义不变子空间，(I) 存在与唯一性定理,计算数学，**2**:1 (1980),1—13.

[7] 孙继广,不变子空间与广义不变子空间，(II) 扰动定理,计算数学,**2**:2(1980), 113—123.

[8] 孙继广,定型矩阵对的特征空间的扰动界限,数学学报,**24**:6(1981),892—903.

[9] 孙继广,一类矩阵对的广义特征值的扰动界限,计算数学，**4**:1(1982),23—29.

[10] 孙继广,广义奇异值的扰动,计算数学，**4**:2(1982)，229—233.

[11] 孙继广，A note on Stewart's theorem for definite matrix pairs, *Lin. Alg.* & *Its Applic.*, **48**(1982), 331—339.

[12] 孙继广，关于 Wielandt-Hoffman 定理,计算数学,**5**:2(1983),208—212.

[13] 孙继广， Perturbation analysis for the generalized eigenvalue and the generalized singular value problem, Lect. Not. in Math. 973, Matrix Pencils, Proc., Pite Havsbad 1982, 221—244(1983).

[14] 孙继广， Orthogonal projections and the perturbation of the eigenvalues of singular pencils, *J. Comp. Math.*, **1**: 1(1983), 63—74.

[15] 孙继广，Perturbation analysis for the generalized singular value problem, *SIAM J. Numer. Anal.*, **20** (1983), 611—625.

[16] 孙继广，The perturbation bounds for eigenspaces of a definite matrix-pair, *Numer. Math.*, **41**(1983), 321—343.

[17] 孙继广，Perturbation theorems for generalized singular values, *J. Comp. Math.*, **1**: 3(1983), 233—242.

[18] 孙继广,关于 Grassmann 流形上的度量(未发表).

[19] 孙继广，关于正规矩阵特征值的扰动,计算数学，**6**:3(1984),334—336.

[20] 孙继广， Estimation of the separation of two matrices, *J. Comp. Math.*, **2**: 3(1984), 189—200.

[21] 孙继广，Estimation of the separation of two matrices. II, *J. Comp. Math.* **3**: 1(1985), 19—26.

[22] 孙继广，Gerschgorin 型定理与奇异束特征值的扰动，计算数学，**7**: 3(1985),

253—264.
[23] 孙继广,正交投影的稳定性,研究生院学报,1:2(1984),123—133.
[24] 何旭初,数值相关性理论及其应用,高等学校计算数学学报,1:1(1979),11—20.
[25] 何旭初,广义逆矩阵的连续性问题——数值相关性理论的应用,高等学校计算数学学报, 1:2(1979), 168—172.
[26] 佟文廷,关于几类矩阵的特征值分布,数学学报, 20: 4(1977), 272—275.
[27] 吴文达,广义逆矩阵,计算机应用与应用数学, 1(1974), 1—25.
[28] 杨曙光, 关于最小二乘问题的多点定向扰动分析,高等学校计算数学学报, 4:3 (1982), 193—201.
[29] 陆启铿, 典型流形与典型域,上海科学技术出版社, 1963.
[30] 曹维潞、郑家栋、谭领、张磊、邓健新, 石钟慈、曹志浩,代数特征值问题(讲义), 1979 年全国数值代数学术讨论会,厦门.
[31] 黄鸿慈,关于矩阵条件数的界,科学通报, 25:18 (1980), 862.
[32] 蒋尔雄,对称正定矩阵 P 条件数的改善,复旦大学学报(自然科学), 9:1(1963), 1—7.
[33] 蒋尔雄, On spectral variation of a nonnormal matrix, *Lin. Alg. and Its Applic.* **42** (1982), 223—241.
[34] 游兆永, 矩阵谱半径的分块估计法, 高等学校计算数学学报, 1:1 (1979), 129—130.
[35] 谭领, 斜映射矩阵和 Householder 变换的推广, 计算数学, 3:1(1981), 66—71.
[36] 樊畿 (Fan Ky), On a theorem of Weyl concerning eigenvalues of linear transformations, *Proc. Nat. Acad. Sci. U. S. A.*, **35** (1949), 652—655.
[37] 樊畿(Fan Ky), Maximum properties and inequalities for the eigenvalues of completely continuous operators, *Proc. Nat. Acad. Sci. U. S. A.*, **37** (1951), 760—766.
[38] 樊畿 (Fan Ky), Note on circular disks containing the eigenvalues of a matrix, *Duke Math. J.*, **25** (1958), 441—445.
[39] Amir Moéz, A. R., Extreme properties of eigenvalues of a Hermitian transformation and singular values of the sum and product of linear transformations, *Duke Math. J.*, **23** (1956), 463—476.
[40] Anderson, A. and Best, G. G., A Gerschgorin-Rayleigh inequality for the eigenvalues of Hermitian matrices, *Lin. and Multilin. Alg.*, 6 (1978), 219—222.
[41] Barnes, E. R. and Hoffman, A. J., On bounds for eigenvalues of real symmetric matrices, *Lin. Alg. and Its Applic.*, 40 (1981), 217—223.
[42] Bart, H., Gohberg, I. and Kaashoek, M. A., Stable factorizations of monic matrix polynomials and stable invariant subspaces, *Integral Equations and Operator Theory*, **1** (1978), 496—517.
[43] Bauer, F. L., A further generalization of the Kontorovic inequality, *Numer. Math.*, 3 (1961), 117—119.
[44] Bauer, F. L., On a field of values subordinate to a norm. *Numer. Math.*, **4** (1962), 103—113.
[45] Bauer, F. L., Optimally scaled matrices, *Numer. Math.*, **5** (1963), 73—

87.
[46] Bauer, F. L., Remarks on optimally scaled matrices, *Numer. Math.*, **13** (1969), 1—3.
[47] Bauer, F. L., Positivity and norms, *Comm. ACM*, **18** (1975), 9—13.
[48] Bauer, F. L., and Fike, C. T., Norms and exclusion theorems, *Numer. Math.*, **2** (1960), 137—141.
[49] Bauer, F. L. and Householder, A. S., Some inequalities involving the Euclidean condition of a matrix, *Numer. Math.*, **2** (1960), 308—311.
[50] Bauer, F. L. and Householder, A. S., Absolute norms and characteristic roots, *Numer. Math.*, **3** (1961), 241—246.
[51] Bauer, F. L., Stoer, J. and Witzgall, C., Absolute and monotonic norms, *Numer. Math.*, **3**(1961), 257—264.
[52] Beckenbach, E. F. and Bellman, R., Inequalities, Springer-Verlag, 1961.
[53] Ben-Israel, A., On error bounds for generalized inverses, *SIAM J. Numer. Anal.*, **3** (1966), 585—592.
[54] Ben-Israel, A., On the geometry of subspaces in Euclidean n-spaces, *SIAM J. Appl. Math.*, **15** (1967), 1184—1198.
[55] Ben-Israel, A. and Greville, T., Generalized Inverses: Theory and Applications, John Wiley, New York, 1974.
[56] Berge, C., Topological Spaces, Oliver and Boyd, 1963.
[57] Bhatia, R., Analysis of spectral variation and some inequalities, *Trans. Amer. Math. Soc.*, **272** (1982), 323—331.
[58] Bhatia, R., Davis, Ch. and McIntosh, A., Perturbation of spectral subspaces and solution of linear operator equations, *Lin. Alg. and Its Applic.*, **52/53** (1983), 45—67.
[59] Bhatia, R. and Friedland, S., Variation of Grassmann powers and spectra, *Lin. Alg. and Its Applic.*, **40** (1981), 1—18.
[60] Bhatia R. and Mukherjea, K., On the rate of change of spectra of operators, *Lin. Alg. and Its Applic.*, **27** (1979), 147—157.
[61] Björk, Å. and Golub, G. H., Numerical methods for computing angles between linear subspaces, *Math. Comp.*, **27** (1973), 579—594.
[62] Brauer, A. and Centry, I. C., Bounds for the greatest characteristic root of an irreducible nonnegative matrix, *Lin. Alg. and Its Applic.*, **8** (1974), 105—107, **13** (1976), 109—114.
[63] Brauer, A. and Mewborn, A. C., The greatest distance between two characteristic roots of a matrix, *Duke Math. J.*, **26** (1959), 653—661.
[64] Brickman, L., On the field of values of matrix, *Proc. Amer. Math. Soc.*, **12** (1961), 61—66.
[65] Brualdi, R. A., Matrices, eigenvalues, and directed graphs, *Lin. and Multilin. Alg.*, **11** (1982), 143—165.
[66] Chatelin-Laborde, F., Perturbation d'une matrice hermitienne ou normale, *Numer. Math.*, **17** (1971), 318—337.
[67] Cline, A. K., Moler, C. B., Stewart, G. W. and Wilkinson, J. H., An estemate for the condition number of a matrix, *SIAM J. Numer. Anal.*, **16** (1979), 368—375.

[68] Cohen, J. E., Convexity of the dominant eigenvalue of an essentially nonnegative matrix, *Proc. Amer. Math. Soc.*, **81** (1981), 657—658.
[69] Crawford, C. R., Reduction of a band-symmetric generalized eigenvalue problem, *Comm. ACM*, **16** (1973), 41—44.
[70] Crawford, C. R., A stable generalized eigenvalue problem, *SIAM J. Numer. Anal.*, **8** (1976), 854—860.
[71] Davis, Ch., Separation of two linear subspaces, *Acta Sci. Math. Szeged*, **19** (1958), 172—187.
[72] Davis, Ch., The rotation of eigenvectors by a perturbation, *J. Math. Anal. Appl.*, **6** (1963), 159—173.
[73] Davis, Ch., The rotation of eigenvectors by a perturbation. II, *J. Math. Anal Appl.*, **11** (1965), 20—27.
[74] Davis, Ch., The Toeplitz-Hausdorff theorem explained, *Canad. Math. Bull.*, **14** (1971), 245—246.
[75] Davis, Ch. and Kahan, W. M., The rotation of eigenvectors by a perturbation. III, *SIAM J. Numer. Anal.*, **7** (1970), 1—46.
[76] Davis, Ch., Kahan, W. M. and Weinberger, H. F., Norm-preserving dilations and their applications to optimal error bounds, *SIAM J. Numer. Anal.*, **19** (1982), 445—469.
[77] Demmel, J., The condition number of equivalence transformations that block diagonalize matrix pencils, *SIAM J. Numer. Anal.*, **20** (1983), 599—610.
[78] Deutsch, E., Matricial norms, *Numer. Math.*, **16** (1970), 73—84.
[79] Deutsch, E., On vectorial norms and pseudonorms, *Proc. Amer. Math. Soc.*, **28** (1971), 18—24.
[80] Deutsch, E., On matricial norms subordinate to vectorial norms, *Math. Z.*, **122** (1971), 142—150.
[81] Deutsch, E., On certain generalized matrix norms, *Lin. Alg. and Its Applic.*, **29** (1980), 113—130.
[82] Dixon, J. D., Estimating extremal eigenvalues and condition numbers of matrices, *SIAM J. Numer. Anal.*, **20** (1983), 812—814.
[83] Eberlein, P. J., On measures of non-normality for matrices, *Amer. Math. Monthly*, **72** (1965), 995—996.
[84] Eldén, L., Perturbation theory for the least squares problem with linear equality constraints, *SIAM J. Numer. Anal.*, **17** (1980), 338—350.
[85] Eldén, L., A weighted pseudoinverse, generalized singular values, and constrained least squares problems, *BIT*, **22** (1982), 487—502.
[86] Elsner, L., On the variation of the spectra of matrices, *Lin. Alg. and Its Applic.*, **47** (1982), 127—138.
[87] Elsner, L. and Sun, J. G. (孙继广), Perturbation theorems for the generalized eigenvalue problem, *Lin. Alg. and Its Applic.*, 48 (1982), 341—357.
[88] Engel, G. M. and Varga, R. S., Minimal Gerschgorin's sets and ω-matrices, *Lin. and Multilin. Alg.*, **5** (1979), 1—10.
[89] Feingold, D. and Varga, R. S., Block diagonally dominant matrices and

generalizations of the Gerschgorin circle theorem, *Pac. J. Math.*, **12** (1962), 1241—1250.

[90] Fix, G. and Heiberger, R., An algorithm for the ill-conditioned generalized eigenvalue problem, *SIAM J. Numer. Anal.*, **9** (1972), 78—88.

[91] Forsythe, G. E. and Strauss, E. G., On best conditioned matrices, *Proc. Amer. Math. Soc.*, **6** (1955), 340—345.

[92] Friedland, S., Convex spectral functions, *Lin. and Multilin. Alg.*, **9** (1981), 299—316.

[93] Gander, W., Least squares with a quadratic constraint, *Numer. Math.*, 36 (1981), 291—307.

[94] Geurts, A. J., A contribution to the theory of condition, *Numer. Math.*, 39 (1982), 85—96.

[95] Gohberg, M., On certain finite dimmensional numerical range and numerical radii, *Lin. and Multilin. Alg.*, **7** (1979), 329—342.

[96] Gohberg, I. C. and Krein, M. G., Introduction to the Theory of Nonselfadjoint Operators, *Amer. Math. Soc. Prov., R. I.* 1969.

[97] Golub, G. H. and Pereyra, V., The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate, *SIAM J. Numer. Anal.*, **10** (1973), 413—432.

[98] Golub, G. H. and Van Loan, C., An analysis of the total least squares problem, *SIAM J. Numer. Anal.*, **17** (1980), 883—893.

[99] Hager, W. W. and Pederson, R. N., Perturbation in eigenvalues, *Lin. Alg. and Its Applic.*, 42 (1982), 39—55.

[100] Hald, O. H., A converse to the Bauer-Fike theorem, *Lin. Alg. and Its Applic.*, **9** (1974), 267—274.

[101] Halmos, P. R., A Hilbert Space Problem Book, Van Nostrand, New York, 1967.

[102] Henrici, P., Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices, *Numer. Math.*, **4** (1962), 24—39.

[103] Hoffman, A. J., On the nonsingularity of real matrices, *Math. Comp.*, **19** (1965), 56—61.

[104] Hoffman, A. J. and Wielandt, H. W., The variation of the spectrum of a normal matrix, *Duke Math. J.*, **20** (1953), 37—39.

[105] Householder, A. S., The approximate solution of matrix problems, *J. Assoc. Comput. Mach.*, **5** (1958), 204—243.

[106] Householder, A. S., The Theory of Matrices in Numerical Analysis, Blaisedell, New York, 1964.

[107] Householder, A. S., Localization of the characteristic roots of matrices, Recent Advances in Matrix Theory (H. Schneider, Ed.), Uni. of Wis. Press, Madison, 1964, 39—60.

[108] Householder, A. S., Separation theorems for normalizable matrices, *Numer. Math.*, **9** (1966), 46—50.

[109] Householder, A. S., Moments and characteristic roots. II, *Numer. Math.*, **11** (1968), 126—128.

[110] Isaacson, E. and Keller, H. B., Analysis of Numerical Methods, John

Wiley, New York, 1966.

[111] Johnson, C. R., Normality and numerical range, *Lin. Alg. and Its Applic.*, **15** (1976), 89—94.

[112] Johnson, C. R., Multiplicativity and compatibility of generalized matrix norms, *Lin. Alg. and Its Applic.*, **16** (1977), 25—37.

[113] Johnston, R L., Gerschgorin theorems for partitioned matrices, *Lin. Alg. and Its Applic.*, **4** (1971), 205—220.

[114] Johnston, R. L., On the computation of inclusion regions for partitioned matrices, *Numer. Math.*, **19** (1972), 238—247.

[115] Kahan, W. M., Spectra of nearly Hermitian matrices, *Proc. Amer. Math. Soc.*, **48** (1975), 11—17.

[116] Kahan, W. M., Parlett, B. N. and Jiang, E. (蒋尔雄), Residual bounds on approximate eigensystems of nonnormal matrices, *SIAM J. Numer. Anal.*, **19** (1982), 470—484.

[117] Kato, T., Perturbation Theory for Linear Operators, Springer-Verlag, New York, 1966.

[118] Kato, T., A Short Introduction to Perturbation Theory for Linear Operators, Springer-Verlag, New York, 1982.

[119] Kress, R., de Vries, H. L. and Wegmann, R., On nonnormal matrices, *Lin. Alg. and Its Applic.*, **8** (1974), 109—120.

[120] Lancaster, P., Explicit solution of linear matrix equations, *SIAM Rev.*, **12** (1970), 544—566.

[121] Lawson, C. L. and Hanson, R. J., Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, 1974.

[122] Marcus, M. and Minc, H., A Survey of Matrix Theory and Matrix Inequalities, Allyn and Bacon, Boston, 1964.

[123] McCarthy, Ch. and Strang, C., Optimal conditioning of matrices, *SIAM J. Numer. Anal.*, **10** (1973), 370—388.

[124] Medley, H. I. and Varga, R. S., On smallest isolated disks for eigenvalues II, III, *Numer. Math.*, **11** (1968), 320—323, 361—369.

[125] Meyer, D. and Veselić, K., On some new inclusion theorems for the eigenvalues of partitioned matrices, *Numer. Math.*, **34** (1980), 431—437.

[126] Mirsky, L., Symmetric gauge functions and unitarily invariant norms, *Quart. J. Math. Oxford*, **11** (1960), 50—59.

[127] Moler, C. B. and Stewart, G. W., An algorithm for the generalized matrix eigenvalue problem, *SIAM J. Numer. Anal.*, **10** (1973), 241—256.

[128] von Neumann, J., Some matrix-inequalities and metrization of matrix-space, Bull. Inst. Math. Mécan. *Univ. Kouybycheff Tomsk*, **1** (1935—1937), 286—300.

[129] O'leary, D. P., Estimating matrix condition numbers, *SIAM J. Sci. Stat. Comput.*, **1** (1980), 205—209.

[130] Ostrowski, A. M., Über die Stetigkeit von charakteristischen Wurzeln in Abhanigkeit von den Matrizenelementen, *Jahresber. Deutsch Math.-Verein.*, **60** (1957), 40—42.

[131] Ostrowski, A. M., Solution of Equations in Euclidean and Banach Spaces,

Academic, 1973.
[132] Paige, C. C., Eigenvalues of perturbed Hermitian matrices, *Lin. Alg. and Its Applic.*, **8** (1974), 1—10.
[133] Paige, C. C., Computer solution and perturbation analysis of generalized linear least squares problems, *Math. Comp.*, **33** (1979), 171—183.
[134] Paige, C. C., A note on a result of Sun Ji-guang: Sensitivity of the CS and GSV decompositions, *SIAM J. Numer. Anal.*, **21** (1984), 186—191.
[135] Paige, C. C. and Saunders, M. A., Towards a generalized singular value decomposition, *SIAM J. Numer. Anal.*, **18** (1981), 398—405.
[136] Parlett, B. N., The Symmetric Eigenvalue Problem, Prentice-Hall, Englewood Cliffs, N. J., 1980.
[137] Parthasarathy, K. R. Eigenvalues of matrix-valued analytic maps, *J. Austral. Math. Soc. Ser. A*, **26** (1978), 179—197.
[138] Pereyra, V., Stability of general systems of linear equations, *Aequat. Math.*, **2** (1969), 194—206.
[139] Peters, G. and Wilkinson, J. H., $Ax=\lambda Bx$ and the generalized eigenproblem, *SIAM J. Numer. Anal.*, **7** (1970), 479—492.
[140] Rice, J. R., A theory of condition, *SIAM J. Numer. Anal.*, **3** (1966), 287—310.
[141] Ruhe, A., Properties of amatrix with a very ill-conditioned eigenproblem, *Numer. Math.*, **15** (1970), 57—60.
[142] Ruhe, A., Perturbation bounds for means of eigenvalues and invariant subspaces, *BIT*, **10** (1970), 343—354.
[143] Ruhe, A., On the closeness of eigenvalues and singular values for almost normal matrices, *Lin. Alg. and Its Applic.*, **11** (1975), 87—94.
[144] Saunders, D. B. and Schneider, H., A symmetric numerical range for matrices, *Numer. Math.*, **26** (1976), 99—105.
[145] Schatten, R., Norm Ideals of Completely Continuous Operators, Springer, Berlin, 1960.
[146] Scheffold, E., Eine Abschätzung für die subdominanten Eigenwerte nichtnegativer Matrizen, *Lin. Alg. and Its Applic.*, **19** (1978), 91—93.
[147] Schneider, H., Regions of exclusion for the latent roots of a matrix *Proc. Amer. Math. Soc.*, **5** (1954), 320—322.
[148] Schöhage, A., Arbitrary perturbations of Hermitian matrices, *Lin. Alg. and Its Applic.*, **24** (1979), 143—149.
[149] Shapiro, A., Optimally scaled matrices, necessary and sufficient conditions, *Numer. Math.*, **39** (1982), 239—245.
[150] Smith, R. A., The condition numbers of the matrix eigenvalue problem, *Numer. Math.*, **10** (1967), 232—240.
[151] Stewart, G. W., On the continuity of the generalized inverse, *SIAM J. Appl. Math.*, **17** (1969), 33—45.
[152] Stewart, G. W., Error bounds for approximate invariant subspaces of closed linear operators, *SIAM J. Numer. Anal.*, **8** (1971), 796—808.
[153] Stewart, G. W., On the sensitivity of the eigenvalue problem $Ax=\lambda Bx$, *SIAM J. Numer. Anal.*, **9** (1972), 669—686.

[154] Stewart, G. W., Intriduction to Matrix Computations, Academic Press, New York, 1973.
[155] Stewart, G. W., Error and perturbation bounds for subspaces associated with certain eigenvalue problems, *SIAM Rev.*, **15** (1973), 727—764.
[156] Stewart, G. W., Gerschgorin theory for the generalized eigenvalue problem $Ax = \lambda Bx$. *Math. Comp.*, 29 (1975), 600—606.
[157] Stewart. G. W., On the perturbation of pseudoinverses, projections and linear squares problems, *SIAM Rev.*, **19** (1977), 634—662.
[158] Stewart G. W., Perturbation theory for the generalized eigenvalue problem, Recent Advances in Numerical Analysis (Proc. Sympos., Math. Res. Center, Univ. Wis., Madison, Wis., 1978) 193—206.
[159] Stewart, G. W., Perturbation bounds for the definite generalized eigenvalue problem, *Lin. Alg. and Its Applic.*, **23** (1979), 69—83.
[160] Stewart, G. W., A note on the perturbation of singular values, *Lin. Alg. and Its Applic.*, **28** (1979), 213—216.
[161] Stewart, G. W., A method for computing the generalized singular value decomposition, Lect. Not. in Math. 973, Matrix Pencils, Proc. Pite Havsbad 1982, 207—220 (1983).
[162] Stewart, G. W., A second order perturbation expansion for small singular values, *Lin. Alg and Its Applic.*, 56 (1984), 231—235.
[163] Stoer, J., On the characterization of least upper bound norms in matrix space, *Numer. Math.*, 6 (1964), 302—314.
[164] Stone, B. J., Best possible ratios of certain matrix norms, *Numer. Math.*, **4** (1962), 114—116.
[165] Sunder, V. S., Distance between normal operators, *Proc. Amer. Math. Soc.*, **84** (1982), 483—484.
[166] Swanson, C. A., An inequality for linear transformations with eigenvalues, *Bull. Amer. Math. Soc.*, 67 (1961), 607—608.
[167] Taussky, O., Eigenvalues of finite matrices: Some topics concerning bounds for eigenvalues of finite matrices, Survey on Numerical Analysis (ed. by J. Todd), McGraw-Hill, New York, 1962, 279—313.
[168] Uhlig, F., A recurring theorem about pairs of quadratic forms and extensions: A survey, *Lin. Alg. and Its Applic.*, **25** (1979), 219—237.
[169] Van der Sluis, A., Condition numbers of the matrix eigenvalue problem, *Numer. Math.*, **10** (1967), 232—240.
[170] Van der Sluis, A., Condition numbers and equilibration of matrices, *Numer. Math.*, **14** (1969), 14—23.
[171] Van der Sluis, A., Stability of solutions of linear algebraic systems, *Numer. Math.*, **14** (1970), 246—251.
[172] Van der Sluis, A., Stability of the solutions of linear least equares problems, *Numer. Math.*, **23** (1975), 241—254.
[173] Van Dooren, P., The computation of Kronecker's canonical form of a singular pencil, *Lin. Alg. and Its Applic.*, **27** (1979), 103—140.
[174] Van Dooren, P., Reducing subspaces-definitions, properties and algorithms, Lect. Not. in Math. 973, Matrix Pencils, Proc., Pite Havsbad

1982, 58—73, (1983).

[175] Van Loan, Ch. F., Generalizing the singular value decomposition, *SIAM J. Numer. Anal.*, **13** (1976), 76—83.

[176] Varah, J. M., Computing invariant subspaces of a general matrix when the eigensystem is poorly conditioned, *Math. Comp.*, **24** (1970), 137—149.

[177] Varah, J. M., Invariant subspace perturbations for a non-normal matrix, IFIP 71 Proceedings, North-Holland, Amsterdam, 1971, 1251—1253.

[178] Varah, J. M., On the separation of two matrices, *SIAM J. Numer. Anal.*, **16** (1979), 216—222.

[179] Varga, R. S., On smallest isolated Gerschgorin disks for eigenvalues, *Numer. Math.*, **6** (1964), 366—376.

[180] Varga, R. S., Minimal Gerschgorin sets, *Pac. J. Math.*, **15** (1965), 719—729.

[181] Varga, R. S., Minimal Gerschgorin sets for partitioned matrices, *SIAM J. Numer. Anal.*, **1** (1971), 493—507.

[182] Wedin, P.-Å., Perturbation bounds in connection with singular value decomposition, *BIT*, **12** (1972), 99—111.

[183] Wedin, P.-Å., Perturbation theory for pseudoinverses, *BIT*, **13** (1973), 217—232.

[184] Wedin, P.-Å., On angles between subspaces of a finite dimmensional inner product space, Lect. Not. in Math. 973, Matrix Pencils, Proc. Pite Havsbad 1982, 263—285 (1983).

[185] Weyl, H., Das asymptotische Verteilungsgesetz der Eigenwerte linear partielles Differentialgleichungen, *Math. Ann.*, **71** (1912), 441—479.

[186] Wielandt, H., Die Einschliessung von Eigenwerten normaler Matrizen, *Math. Ann.*, **121** (1949), 234—241.

[187] Wielandt, H., An extremum property of sums of eigenvalues, *Proc. Amer. Math. Soc.*, 6 (1955), 106—110.

[188] Wielandt, H., On eigenvalues of sums of normal matrices, *Pac. J. Math.*, **5** (1955), 633—638.

[189] Wilkinson, J. H., The Algebraic Eigenvalue Problem, Charendon Press, Oxford, 1965.

[190] Wilkinson, J. H., Almost diagonal matrices with multiple or close eigenvalues, *Lin. Alg. and Its Applic.*, **1** (1968), 1—12.

[191] Wilkinson, J. H., Note on matrices with a very ill-conditioned eigenproblem, *Numer. Math.*, **19** (1972), 176—178.

[192] Wilkinson, J. H., Kronecker's canonical form and the QZ algorithm, *Lin. Alg. and Its Applic.*, **28** (1979), 285—303.

[193] Wilkinson, J. H. and Golub, G., Ill-conditioned eigensystems and the computation of the Jordan canonical form, *SIAM Rev.*, **18** (1976), 578—619.

[194] Williamson, J., Note on a principal axis transformation for non-Hermitian matrices, *Bull. Amer. Math. Soc.*, **45** (1939), 920—922.

[195] Wolkowicz, H. and Styan, G. P. H., Bounds for eigenvalues using traces, *Lin. Alg. and Its Applic.*, **29** (1980), 471—506.

[196] Wolkowicz, H. and Styan, G. P. H., More bounds for eigenvalues using traces, *Lin. Alg. and Its Applic.*, **31** (1980), 1—17.

[197] Гантмахер, Ф. Р., Теория матриц, Издательство «Наука», ФМЛ, Москва, 1966.

[198] Лидский, В. Б., О собственных значениях суммы и произведения симметрических матриц, *ДАН*, **75**(1950), 769—772.